

Schreiber, David

135252

From: Ramirez, Delia
Sent: Thursday, October 14, 2004 1:10 PM
To: Schreiber, David
Subject: 10/041018

Hi,

I would like to request the following searches:

1. seq id 1 in the nucleic acid databases (commercial and interference)
2. seq id 361 in the nucleic acid databases (commercial and interference)
3. seq id 383 in the protein databases (commercial and interference)
4. seq id 22 in the protein databases (commercial and interference)
5. alignment of seq id 22 and seq id 1
6. alignment of seq id 383 and seq id 361

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

30

SEARCH REQUEST FORM**Scientific and Technical Information Center**

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

| | Type of Search | Vendors and cost where applicable |
|--|--------------------------|-----------------------------------|
| Searcher: <u>D. Schreiber</u> | NA Sequence (#) <u>4</u> | STN _____ |
| Searcher Phone #: <u>272-2526</u> | AA Sequence (#) <u>4</u> | Dialog _____ |
| Searcher Location: <u>Rensselaer E01 #61</u> | Structure (#) _____ | Questel/Orbit _____ |
| Date Searcher Picked Up: _____ | Bibliographic _____ | Dr.Link _____ |
| Date Completed: <u>10/18</u> | Litigation _____ | Lexis/Nexis _____ |
| Searcher Prep & Review Time: <u>15</u> | Fulltext _____ | Sequence Systems <u>CompuGen</u> |
| Clerical Prep Time: _____ | Patent Family _____ | WWW/Internet _____ |
| Online Time: <u>27</u> | Other _____ | Other (specify) _____ |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:30:51 ; Search time 292.161 Seconds
(without alignments)
1543.991 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSCIASPLITKSNRPA.....TVKDIYNPLVNVNEBQR 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|------------------------|
| 1 | 4113 | 100.0 | 784 | 2 Q9XEI0 | Q9xei0 stevia reba |
| 2 | 4026 | 97.9 | 784 | 2 Q9XEI9 | Q9xeh9 stevia reba |
| 3 | 2855.5 | 69.4 | 788 | 2 Q9FXV8 | Q9fxv8 lactuca sat |
| 4 | 2101.5 | 51.1 | 785 | 2 Q9PRX5 | Q9prx5 cucumis sat |
| 5 | 2023.5 | 49.2 | 789 | 2 Q93548 | Q93548 cucurbita m |
| 6 | 2021 | 49.1 | 785 | 2 Q64952 | Q64952 arabidopsis |
| 7 | 1556.5 | 37.8 | 683 | 2 Q7XLE0 | Q7xle0 oryza sativ |
| 8 | 1530 | 37.2 | 821 | 2 Q6Z5U6 | Q6z5u6 oryza sativ |
| 9 | 1530 | 37.2 | 821 | 2 BAD17270 | Bad17270 oryza sat |
| 10 | 1530 | 37.2 | 830 | 2 Q85282 | Q85282 oryza sativ |
| 11 | 1525.5 | 37.1 | 802 | 2 Q8VY94 | Q8vy94 oryza sativ |
| 12 | 1525.5 | 37.1 | 802 | 2 BAD17672 | Bad17672 oryza sat |
| 13 | 1319 | 32.1 | 770 | 2 Q7XK63 | Q7xk63 oryza sativ |
| 14 | 1312 | 31.9 | 913 | 2 Q7XLD9 | Q7xld9 oryza sativ |
| 15 | 1276 | 31.0 | 730 | 2 Q7XLE2 | Q7xle2 oryza sativ |
| 16 | 1144.5 | 27.8 | 673 | 2 Q6Z5U0 | Q6z5u0 oryza sativ |
| 17 | 1144.5 | 27.8 | 673 | 2 BAD17276 | Bad17276 oryza sat |
| 18 | 1102 | 26.8 | 590 | 2 Q84ZW8 | Q84zw8 zea mays (m |
| 19 | 1075 | 26.1 | 873 | 2 Q947C4 | Q947c4 ginkgo bilo |
| 20 | 1075 | 26.1 | 873 | 2 AAS89668 | Aas89668 ginkgo bi |
| 21 | 1062.5 | 25.8 | 868 | 2 Q38710 | Q38710 abies grand |
| 22 | 1062.5 | 25.6 | 853 | 2 Q94FW1 | Q94fw1 abies grand |
| 23 | 1017.5 | 24.7 | 862 | 1 TASY_TAXBA | Tasy_taxba taxus bac |
| 24 | 1017.5 | 24.7 | 862 | 2 Q6TBY0 | Q6tby0 taxus bacca |
| 25 | 1017.5 | 24.7 | 862 | 2 AAR02861 | Aar02861 taxus bac |
| 26 | 1017.5 | 24.7 | 862 | 2 Q6SAG0 | Q6sag0 taxus x med |
| 27 | 1016.5 | 24.7 | 862 | 2 Q6SAG0 | Q6sag0 taxus x med |
| 28 | 1016.5 | 24.7 | 862 | 2 AAS18603 | Aas18603 taxus x m |
| 29 | 1013.5 | 24.6 | 862 | 1 TASY_TAXBR | Tasy_taxbr taxus brevi |
| 30 | 990.5 | 24.1 | 862 | 1 TASY_TAXCH | Tasy_taxch taxus chine |
| 31 | 965 | 23.5 | 782 | 2 Q9SAU6 | Q9sau6 abies grand |

| | | | | | | |
|----|-------|------|-----|---|----------|--------------------|
| 32 | 965 | 23.5 | 817 | 2 | O81086 | O81086 abies grand |
| 33 | 962 | 23.4 | 816 | 2 | Q94FW2 | Q94fw2 abies grand |
| 34 | 925 | 22.5 | 458 | 2 | Q9XEI5 | Q9xei5 zea mays (m |
| 35 | 876.5 | 21.3 | 871 | 2 | Q92PN5 | Q92pn5 clarkia con |
| 36 | 863 | 21.0 | 870 | 2 | Q96376 | Q96376 clarkia bre |
| 37 | 853 | 20.7 | 877 | 2 | Q93YV0 | Q93yv0 arabidopsis |
| 38 | 837.5 | 20.4 | 813 | 2 | Q92PN4 | Q92pn4 clarkia bre |
| 39 | 837.5 | 20.4 | 878 | 2 | Q84U03 | Q84u03 arabidopsis |
| 40 | 813.5 | 19.8 | 870 | 2 | O22733 | O22733 arabidopsis |
| 41 | 760 | 18.5 | 829 | 2 | Q9MAX2 | Q9max2 croton subl |
| 42 | 732 | 17.8 | 787 | 2 | O22667 | O22667 stevia reba |
| 43 | 725.5 | 17.6 | 801 | 2 | O04408 | O04408 pisum sativ |
| 44 | 723.5 | 17.6 | 479 | 2 | Q6VTF8 | Q6vtf8 oryza sativ |
| 45 | 723.5 | 17.6 | 479 | 2 | BAD17771 | Bad17771 oryza sat |

ALIGNMENTS

RESULT 1

| | | | |
|--------|---|------|---------|
| Q9XEI0 | PRELIMINARY; | PRT; | 784 AA. |
| ID | Q9XEI0 | | |
| AC | Q9XEI0; | | |
| DT | 01-NOV-1999 (TREMELrel. 12, Created) | | |
| DT | 01-NOV-1999 (TREMELrel. 12, Last sequence update) | | |
| DT | 01-OCT-2003 (TREMELrel. 25, Last annotation update) | | |
| DE | kaurene synthase. | | |
| GN | Name=KS22-1; | | |
| OS | Stevia rebaudiana (Stevia). | | |
| OC | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | |
| OC | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; | | |
| OC | campanulids; Asterales; Asteraceae; Eupatorieae; Eupatoriaceae; Stevia. | | |
| OX | NCBI_taxid=55670; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=Leaf; | | |
| RX | MEDLINE=99435880; PubMed=10504563; | | |
| RA | Richman A.S., Gijzen M., Starratt A.N., Yang Z., Brandle J.E.; | | |
| RT | "Diterpene synthesis in Stevia rebaudiana: recruitment and up- | | |
| RT | regulation of key enzymes from the gibberellin biosynthetic pathway."; | | |
| RL | Plant J. 19:411-421(1999). | | |
| DR | EMBL; AF097311; AAD34295.1; -- | | |
| DR | HSSP; Q40577; 1HX9. | | |
| DR | GO; GO:0016829; F:lyase activity; IEA. | | |
| DR | GO; GO:0008152; P:metabolism; IEA. | | |
| DR | InterPro; IPR005630; Terpene synth C. | | |
| DR | InterPro; IPR008949; Terpenoid synth. | | |
| DR | InterPro; IPR008930; Terp cyc Toroid. | | |
| DR | InterPro; IPR001906; Terp synth-like. | | |
| DR | Pfam; PF01397; Terpene synth_1. | | |
| DR | Pfam; PF03936; Terpene synth_C; 1. | | |
| SQ | SEQUENCE 784 AA; 89400 MW; 284BD3972DC09051 CRC64; | | |

| | | | | | | | | |
|-----------------------|-----------------|---------------------|--------------|-----------------|-----------------------|--------------------|------------------|-----|
| Query Match | 100.0%; | Score 4113; | DB 2; | Length 784; | | | | |
| Best Local Similarity | 100.0%; | Pred. No. 4.1e-234; | | | | | | |
| Matches 784; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; | | | | |
| QY | 1 | MNLSLCIASP | LLTKSNRPAALS | AIHTASTSHGGQTNP | TNLIIDTTKRIQKQFNVEISV | 60 | | |
| Db | 1 | MNLSLCIASP | LLTKSNRPAALS | AIHTASTSHGGQTNP | TNLIIDTTKRIQKQFNVEISV | 60 | | |
| QY | 61 | SSYDTAWAV | MPSPNSPKS | PCFP | ECLNWLNNQLNDG | SGVLNHNHPLPKDSLSST | 120 | |
| Db | 61 | SSYDTAWAV | MPSPNSPKS | PCFP | ECLNWLNNQLNDG | SGVLNHNHPLPKDSLSST | 120 | |
| QY | 121 | LACIVALKRN | VNVEDQ | INKGLSF | IESNLASATE | KSQSPSGFDI | IFPGILEYAKNLIDNL | 180 |
| Db | 121 | LACIVALKRN | VNVEDQ | INKGLSF | IESNLASATE | KSQSPSGFDI | IFPGILEYAKNLIDNL | 180 |
| QY | 181 | LSKQTFDPL | MLHRELE | QKCHSNEMD | GYLAYISE | GLNLYDWN | VKKYQMKNGSVFNSP | 240 |
| Db | 181 | LSKQTFDPL | MLHRELE | QKCHSNEMD | GYLAYISE | GLNLYDWN | VKKYQMKNGSVFNSP | 240 |

241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLIGISHFRVE 300
 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLIGISHFRVE 300
 301 IKNVLDYTRCWRVERDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELAKDEYAA 360
 301 IKNVLDYTRCWRVERDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELAKDEYAA 360
 361 LETYHSHILYQEDLSSGKQILKSAADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 361 LETYHSHILYQEDLSSGKQILKSAADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 421 RINTRENIQLYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 421 RINTRENIQLYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 481 ENKLDQKTAROKTAYCYFSAATLSSPELSDARISWAKNGIILTTVVDDFFDGGTIDEL 540
 481 ENKLDQKTAROKTAYCYFSAATLSSPELSDARISWAKNGIILTTVVDDFFDGGTIDEL 540
 541 TNLIOQVEKWNVDVDCCESEHVRILFLAKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 541 TNLIOQVEKWNVDVDCCESEHVRILFLAKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPPIYFVGPKLSEIIVSSSEYHNL 660
 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPPIYFVGPKLSEIIVSSSEYHNL 660
 661 FKLMSTQGRLLNDIHSFKREFKEGKGLNAVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 661 FKLMSTQGRLLNDIHSFKREFKEGKGLNAVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 721 MKLIFENGSIIVPRACKDAFWNNCHVLNFFYANDDGTGTNTILDFTVKDIIYNPLVLNEN 780
 721 MKLIFENGSIIVPRACKDAFWNNCHVLNFFYANDDGTGTNTILDFTVKDIIYNPLVLNEN 780
 781 EQOR 784
 781 EQOR 784

RESULT 2

Q9XEH9 PRELIMINARY; PRT; 784 AA.
 AC Q9XEH9
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kaurene synthase.
 GN Name=KSL-1;
 OS Stevia rebaudiana (Stevia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC campanulids; Asterales; Asteraceae; Asterioideae; Eupatorioideae; Stevia.
 OX NCBI_TaxID=55670;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=99435880; PubMed=10504563;
 RA Richman A.S., Gijzen M., Starratt A.N., Yang Z., Brandle J.E.;
 RT "piterpene synthesis in Stevia rebaudiana: recruitment and up-
 regulation of key enzymes from the gibberellin biosynthetic pathway.";
 RL Plant J. 19:411-421(1999).
 DR EMBL; AF097310; AAD34294.1; -.
 DR HSSP; O81192; INLZ.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR005530; Terpene synth C.
 DR InterPro; IPR008949; Terpenoid synth.
 DR InterPro; IPR008930; Terp_cyc.teroid.
 DR InterPro; IPR001906; Terp_synth-like.
 DR Pfam; PF01397; Terpene synth; 1.
 DR Pfam; PF03936; Terpene synth_C; 1.

SQ SEQUENCE 784 AA; 89401 MW; D83B95DE21C0B35C CRC64;
 Query Match 97.9%; Score 4026; DB 2; Length 784;
 Best Local Similarity 98.0%; Pred. No. 5.6e-229;
 Matches 768; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MNLSLCIASPLITKSNRPAALSAIHTASTSHGQNTPTNLIIDTTKERTOKOKFNVEISV 60
 Db 1 MNLSLCIASPLITKSNRPAALSAIHTASTSHGQNTPTNLIIDTTKERTOKOKFNVEISV 60
 QY 61 SSIYTAWAMVSPSPKSPCPPECLNMLINNQLDNGWGLVNHHTHHNHPILKOSLSST 120
 Db 61 SSIYTAWAMVSPSPKSPCPPECLNMLINNQLDNGWGLVNHHTHHNHPILKOSLSST 120
 QY 121 LACTIVALKEWVNGEIQNGLSFIENSLASATKESQSPGIFGDIIPFGLLEAKNLDINL 180
 Db 121 LACTIVALKEWVNGEIQNGLSFIENSLASATKESQSPGIFGDIIPFGLLEAKNLDINL 180
 QY 181 LSKQTDFFSLMLHRELEQKCHSNEMDGYLAVISEGLNLYDMNMVVKYOMKNGSVFNSP 240
 Db 181 LSKQTDFFSLMLHRELEQKCHSNEMDGYLAVISEGLNLYDMNMVVKYOMKNGSVFNSP 240
 QY 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLIGISHFRVE 300
 Db 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLIGISHFRVE 300
 QY 301 IKNVLDYTRCWRVERDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELAKDEYAA 360
 Db 301 IKNVLDYTRCWRVERDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELAKDEYAA 360
 QY 361 LETYHSHILYQEDLSSGKQILKSAADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 Db 361 LETYHSHILYQEDLSSGKQILKSAADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 QY 421 RINTRENIQLYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 Db 421 RINTRENIQLYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 QY 481 ENKLDQKTAROKTAYCYFSAATLSSPELSDARISWAKNGIILTTVVDDFFDGGTIDEL 540
 Db 481 ENKLDQKTAROKTAYCYFSAATLSSPELSDARISWAKNGIILTTVVDDFFDGGTIDEL 540
 QY 541 TNLIOQVEKWNVDVDCCESEHVRILFLAKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 Db 541 TNLIOQVEKWNVDVDCCESEHVRILFLAKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 QY 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPPIYFVGPKLSEIIVSSSEYHNL 660
 Db 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPPIYFVGPKLSEIIVSSSEYHNL 660
 QY 661 FKLMSTQGRLLNDIHSFKREFKEGKGLNAVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 Db 661 FKLMSTQGRLLNDIHSFKREFKEGKGLNAVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 QY 721 MKLIFENGSIIVPRACKDAFWNNCHVLNFFYANDDGTGTNTILDFTVKDIIYNPLVLNEN 780
 Db 721 MKLIFENGSIIVPRACKDAFWNNCHVLNFFYANDDGTGTNTILDFTVKDIIYNPLVLNEN 780
 QY 781 EQOR 784
 Db 781 EQOR 784
 RESULT 3
 Q9FXV8 PRELIMINARY; PRT; 788 AA.
 ID Q9FXV8
 AC Q9FXV8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ent-kaurene synthase No1.
 GN Name=LKSL1;
 OS Lactuca sativa (Garden lettuce).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Asterales; Asteraceae; Cichorioideae; Cichoriaceae;
 OC Lactuca.
 OX NCBI_TaxID=4236;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Grand Rapids;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB031205; BAB12441.1; -
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR005630; Terpene synth C.
 DR InterPro; IPR008949; Terpenoid synth.
 DR InterPro; IPR008930; Terp_cyc toroid.
 DR InterPro; IPR001906; Terp_synth-like.
 DR Pfam; PF01397; Terpene synth; 1.
 DR Pfam; PF03936; Terpene synth_C; 1.
 SQ SEQUENCE 788 AA; 89611 MW; 0AA2111132810CE7 CRC64;
 Query Match 69.4%; Score 2855.5; DB 2; Length 788;
 Best Local Similarity 67.8%; Pred. No. 6.4e-160;
 Matches 541; Conservative 102; Mismatches 116; Indels 39; Gaps 6;
 QY 1 MNLSLCIASPLITKSN-----RPAALSAIHTASTSHGGQTNPTNLT 44
 Db 1 MNIAQITSSAMLVPSHTPHRSVVNCCWQYNP---SGLRTAS-SQAGQVPTVMTLDV 56
 QY 45 TKERIQKQKVEISVSYDTAWAVMPSNPKSPKPCPECLNMLNQLNDGSGWLVNH 104
 Db 57 TKERIRKLFNNVEVSVSYDTAWAVMPSNPKSPKPCPECLNMLNQLNDGSGWLVNH 114
 QY 105 THNHNHLLKDSLSTLACIIVALKRWNVGEOINKGLSFIESNLASATEKSPQSPIGDI 164
 Db 115 --PHQSPLIKDTLSTLACVILAKRWNVGEOINKGLHIESNFASVTDKQASPFQDI 172
 QY 165 IFPGLEAYAKNLDNLKASQKDFSLMLHKLRELEKQKCHSNEMDGVIAVISEGLNLYDWN 224
 Db 173 IFPGMLEYAKDLKIPLNQTHLSVMLHERLELRCHNSGREAYIAVISEGLNLDWN 232
 QY 225 MVKKYQMKNGSVNPSATAAAFINHQNGCLNLYNLLDKFGNAVPTVYPHDLFIRLSM 284
 Db 233 MVKKYQMKNGSLFNSPATSALVLIHQVAGCLHLYLSLLDKFGNAVPTVYPIDLYVRLSM 292
 QY 285 VDTLRLGISHFVEIKNVLDETYRCWVERDEQIEMDVVTCALAFRLRLNGEVSDDP 344
 Db 293 VDTLRLGIKHFVVEIQNVLDETYRCWVQGDVQIFMDVVTICALAFRLVRLSNGEVSDDP 352
 QY 345 LAEITN-----ELALKDQYAALETYHSHILYQEDLSGKOLKLSADFLKEIISTDS 396
 Db 353 LAKITKEDYMNSPKPKDYVTSLEVYKASQIYQBELAFREQLNLS-----YLPSS 405
 QY 397 NLSKLHKEVENALKPINTGLERINTRNIOYVNDTRILKTYHSSNISNTDYLRL 456
 Db 406 NKLSNYILKEVDLAKFPFNGSLERMSRTRNIEHYNLNHTRLKTYSSSNISNKVYLK 465
 QY 457 AVEDFYTCQSYREELGLERWVENKLDQKFAKQKTAICYFSAVATLSPPELSDARIS 516
 Db 466 AVQDFNECQSYCEBELKLERWVENKLDQKFAKQKTAICYFSAVATLSPPELSDARIS 525
 QY 517 WAKNGILTVVDDPFDIGTITDELITNLIOCEKKNVDVDDKCCSEHVRILFLALKDAICW 576
 Db 526 WAKSSILTVTIDDFDVGSGNDELNVFHIIEKKNVNVENDCCSEEVGVLFALKDAVCW 585
 QY 577 IGDAFKWQARDVTSHTVITQWLELMSNLRBAIWRDAYPTLINEYMENAYVSPALGPV 636
 Db 586 IGDKAFKIQERNITSHVIEIWLVLKSMLEAIWAKDGSIPTINEYMENGVSFALGPV 645
 QY 637 KPAIFYGPKLSEELVSESSEYHNLFKLMSTQRLNIDHSPKRFKGGKLNVALHISNG 696
 Db 646 LPTLYFLGVKLSSEVVSQSEYHKLIEYVNSTQRLMNDIHSFKREKKAGKLNVALYMSDG 705

QY 697 BSGKVEEVEWREMMVMKIKRKLMLKLIFFRENGSIYPRACKDAFNMCHLVNFFYANDDG 756
 Db 706 KSGSVEEVEWREMMVMKIKRKLMLKLIFFRENGSIYPRACKDAFNMCHLVNFFYANDDG 765
 QY 757 FTGNTILDTVKDIYINPL 774
 Db 766 FTGNAILDVKKEIIEPV 783
 RESULT 4
 Q9FRX5
 ID Q9FRX5 PRELIMINARY; PRT; 785 AA.
 AC Q9FRX5
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Ent-kaurene synthase.
 GN Name=CskS1;
 OS Cucumis sativus (Cucumber).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids 1; Cucurbitales; Cucurbitaceae; Cucumis.
 OX NCBI_TaxID=3659;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shirai K., Fujino K., Masuda K.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB045310; BAB19275.1; -
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR001330; Prenyltrans.
 DR InterPro; IPR005630; Terpene synth C.
 DR InterPro; IPR008949; Terpenoid synth.
 DR InterPro; IPR008930; Terp_cyc toroid.
 DR InterPro; IPR001906; Terp_synth-like.
 DR Pfam; PF00432; Prenyltrans; 1.
 DR Pfam; PF01397; Terpene synth; 1.
 DR Pfam; PF03936; Terpene synth_C; 1.
 SQ SEQUENCE 785 AA; 88706 MW; 8EE26D9D7D33FE350 CRC64;
 Query Match 51.1%; Score 2101.5; DB 2; Length 785;
 Best Local Similarity 50.9%; Pred. No. 1.9e-115;
 Matches 413; Conservative 136; Mismatches 194; Indels 69; Gaps 10;
 QY 1 MNLSLCIASPLITKSNRPAALSAIHTASTSH-----GGQTNPTNLIIDTTKRIQKQ 52
 Db 1 MNLS-----RPTNLGCF-TASSASLFPGLDVGTKTKTGALRPFETKERIKKL 47
 QY 53 FKQVEISVSYDTAWAVMPSNPKSPKPCPECLNMLNQLNDGSGWLVNHNHPL 112
 Db 48 FKQVELSISAYDTAWAVMPSNPKSPKPCPECLNMLNQLNDGSGWLVNHNHPL 102
 QY 113 -LKDLSLSTLACIIVALKRWNVGEOINKGLSFIESNLASATEKSPQSPIGDIIFPGGLE 171
 Db 103 VMKATLSTLACVILAKRWIGDDHMSKALSFIKNSIASATDENQKQSPVGFDIIFPGWIE 162
 QY 172 YAKNLDNLKSKQDFSLMLHKLRELEKQKCHSNEMDGVIAVISEGLNLYDWNVKKY 229
 Db 163 YAKDLNLPLASNNVDALVQKKELELRSCCSNSEGKAVIAYVSEGIKLDQWEMVRY 222
 QY 230 QMKNGSVNPSATAAAFINHQNGCLNLYNLLDKFGNAVPTVYPHDLFRLSMVDIE 289
 Db 223 QMKNGSLFSPSTAVAFMRNDGCFNYLRSVLQKPHSSVPATYIPDILVARLHMVDSLQ 282
 QY 290 RLGISHHFVEIKNVLDETYRCWVERDEQIEMDVVTCALAFRLRLNGEVSDDPDLAET 349
 Db 283 KLGIHGFKDIEIRSLVDETYSCWQGENFLDASTCAMAFLRLVEGYDVSSDQLTQFS 342
 QY 350 NEL-----ALKDEYAALETYHSHILYQED-----LSSGKOLKLS 384
 Db 343 EGLFNSCLGHLKDFASLELFRASQIIYDPDFILENINSWTSRFLNHGLSSG----- 396
 QY 385 ADFLKEITISDTSNLSKLHKEVENALKFPINTGLERINTRNIOYVNDTRILKTYH 444

Db 397 -----SVHSDRIERLVKQAVNAFFPYKSTLERLSNKRALLESYSGDIVRISKATA 448
QY 445 SNIQNTDYLRLAVEDFTYQSIYREBELKGLRWVVENKLDOLKPARQKATAYCYFSAAT 504
Db 449 CLNFGQDPLELAVEDFTNLTGHRKELKELKWIENKLDKLFARQKLAYCYFSAAT 508
QY 505 LSSPELSDARISWAKNGIITVVDDFDIGGTIDRLTNLIQCEKWNVDVDCCESEHYR 564
Db 509 LTPSELCDARLSWAKNGVITVVDDFDVGGSEBELVNLILQVEKWDASGETGYCSKEVE 568
QY 565 ILFLAKDAICWIGDAFKQARDVTSHVIOTWLELMSMLREAIWTRDAYVPTLNEYME 624
Db 569 IIFLALHSITICEIGKALPQGRSVNRNVIDIWLALLESMRKEAEWLKNVPSLDEYME 628
QY 625 NAYVSPALOPTVPAIYFVGPKLSEBIVSESSSEYHNLFKLMSTQGRLLNDIHSFKREFEG 684
Db 629 NGVSPALGPIVLPTLYFVGPKLPEEIVGNCEYKFLFKLMSTGRLLNTRTFDRESSEG 688
QY 685 KLNALHLSNGESGKVEEVEVMMWMIKMKRKLMLKLIFFENGISVPRACKDAFNMWC 744
Db 689 KLNALSLYMWISAGGKLTKEATEAMKGDVDRTRERLLRLVQEN-STIPRACKDLFWKMS 747
QY 745 HVNLFFYANDGFTGNTGNTILDTVVKDIINPLVL 776
Db 748 CVVHLFYRKDDGFTSHELANSKALFEQPMVL 779

RESULT 5
Q39548
ID Q39548 PRELIMINARY; PRT; 789 AA.
AC Q39548;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Ent-kaurene synthase B.
OS Cucurbita maxima (Pumpkin) (winter squash).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid I; Cucurbitales; Cucurbitaceae; Cucurbita.
OX NCBI_TaxID=3661;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Immature seeds;
RX MEDLINE=96367664; PubMed=8771778;
RA Yamauchi S., Saito T., Abe H., Yamane H., Murofushi N., Kamiya Y.;
RT "Molecular cloning and characterization of a cDNA encoding the
RT gibberellin biosynthetic enzyme ent-kaurene synthase B from pumpkin
RT (Cucurbita maxima L.).";
RL Plant J. 10:203-213 (1996).
DR EMBL; U43904; AAB39482.1; -.
DR FTR; T09672; T09672.
DR GO; GO:0016829; F1yase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR009022; EFG III V.
DR InterPro; IPR005630; Terpene synth C.
DR InterPro; IPR008949; Terpenoid synth.
DR InterPro; IPR008930; Terp. cyc. toroid.
DR InterPro; IPR001906; Terp. synth.-like.
DR Pfam; PF01397; Terpene synth; 1.
DR Pfam; PF03936; Terpene synth C; 1.
SQ SEQUENCE 789 AA; 89361 MW; 8CC87F2884CCD7F2 CRC64;

Query Match 49.2%; Score 2023.5; DB 2; Length 789;
Best Local Similarity 50.08; Pred. No. 7.8e-111;
Matches 393; Conservative 150; Mismatches 212; Indels 31; Gaps 8;

QY 16 NRPAALSIAHTASTSGGQNPNTLIIDT-----TKERIQQKFNVEISSYD 64
Db 4 SRTGVARFAASSSSSSSASIFFGVDDVTTTKGALHFEETKRIKFLDKVELSVAYD 63
QY 65 TAWVAVPSPNSPKSPGCPFPCLNWLINQNLNDGSLWLNTHNNHPLL-KDLSSTLAC 123
|||||
|||||
|||||

Db 64 TAWVAVPSPNSLNQPLFPPECINNVLVDSQHADSGWLL-----HNDQLLMKANLSTLAC 118
QY 124 IVALKRWVGEQDINKGLSFESNLASATEKSQSPIGFDIIFPGGLLEYAKNLDINLISK 183
Db 119 VUTLKRNIIGHDMKALDFIKSNIASATDENQSPVGFDIIFPGWIEVAKDNLNPLA 178
QY 184 QTFDSIMLHKRELEQKCRCHSNEMDG--YLAYTSEGNGLYNNWVKYQKNGSVNPS 241
Db 179 PTNVDAVVRKELELRSCRNSSEGGKAYLAYVSEGITKLQDMWVQYQKNGSLFNSPS 238
QY 242 ATAAAFIHNQNGCLNLSLIDKFGNAVPTVYVPHDLFRLSNVDTIERLGHSHPRVEI 301
Db 239 TTAAPAFMRNDDGCFDRLSLQKFDGSPVTIPIIDIVARLHNVDLSQKFGIARHKEI 298
QY 302 KNLVDTRYCWRVERDQIFMDVVTCALAPRILRLINGEYVSPDPAEITNEL-----AL 354
Db 299 RSVLDTRYCWNQGENIFLDASTCAMAFLMRVEGYDVSSDQLTQFSEDFPNCUGGYL 358
QY 355 KDEYALETYHASHILYQEDLSSGKOILK-SADFLKEIISTD---SNRLSKLIHKVEVA 410
Db 359 KDFGASLELYKASQIITHPDESIVLENISWTSRFLKHGLSSDSVMSDRDTSVVKQAVNA 418
QY 411 IKPPINTGLERINTRNIQIXNDVNTIRILKTTVHSSNISNTDYLRLAVEDFTYQSIYRE 470
Db 419 LEFPYNATLERLISKRAMESYSGDIVRISKSPVACLNFGHQDFLELAVEDFTLQRIHLK 478
QY 471 ELKGLRWVVENKLDOLKPARQKATAYCYFSAATLSSPELSDARISWAKNGIITVVDDF 530
Db 479 ELEELQRWVVENKLDLKEFLHLYCYFAAAATLTDPELHDARIAWAQNGVLTVVDDF 538
QY 531 FPIGTTIDTLNLIQCEKWNVDVDCCESEHVRILFLAKDAICWIGDAFKQARDVT 590
Db 539 YDGGSEBELDNLIELVEKWDPDGEVGYCKDVEIVFLALHSTVCEIGRRALVQGRSV 598
QY 591 SHVIOTWLELMSMLREAIWTRDAYVPTLNEYMENAVVSFALGPIYVGPKLSE 650
Db 599 RNVIDGWLALLKVMRKEAENKVPVSMGMEQARVVSFALGPIIPLMFFVGPKLSE 658
QY 651 IVESSEYHNLFKLMSTQGRLLNDIHSFKREFEGKLNALHLSNGESGKVEEVEVMM 710
Db 659 MGSCEYQKLYKLMSTAGRLKNDIRSVDRCKEGLNLSLWMDGGNVTKEATEAIK 718
QY 711 MMKNRKELMKLIFFENGISVPRACKDAFNMCHVNLNFFYANDGFTGNTILDTVKDI 770
Db 719 GDFERAIRELLGLVQEN-TTIPRACKDLFWKLSIVNLFYMEDDGYTSNRLMNTVKAMF 777
QY 771 YNPLVL 776
Db 778 EQMDL 783

RESULT 6
O64952 PRELIMINARY; PRT; 785 AA.
AC O64952;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Ent-kaurene synthase.
GN Name=GA2;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98205064; PubMed=9536043;
RA Yamauchi S., Sun T., Kawaide H., Kamiya Y.;
RT "The GA2 locus of Arabidopsis thaliana encodes ent-kaurene synthase of
RT gibberellin biosynthesis."
RL Plant Physiol. 116:1271-1278 (1998).
DR EMBL; AF034774; AAC39443.1; -.

Db 500 CYFSGAATLSPSPDARISWAKGGVLTIVDDFDVGGSGKELENLHLVKEWDLNGVP 559
 QY 557 DCSSEHVRILFLAKDAICWIGDAEFKQARDVTSVHIQTWLELNMMLREAIWTEDAYV 616
 Db 560 EYSSEHEIIFSVLRDTILETGDKAFYQGRNVTHHVIKVLDDLKMLREAEWSKST 619
 QY 617 PTLNEYMENAYVSFALGPVKPAIYFVGPKLSEIIVESSEYHNLFKLMSTQGRLLNDIHS 676
 Db 620 PSELDYMENAYVSFALGPVLPATYLIIGPLPEKTVDSHQYNQYKLVSTWGLNDIQ 679
 QY 677 FKREFKGNALVALHLSNGESGKVEEVEVMMMKMKRKEMLKLIPEENGSIIVPRAC 736
 Db 680 FKRESAEGKLNVALHMERDNRSKEVIESMKGLAERKRERHKLVLKEKGSVVPREC 739
 QY 737 KDAFWNMCHVLPFFYANDGFTNTILDVTKLIYNPLVLNVEN 780
 Db 740 KEAFLKMSVNLNLFYRKDDGFTSNDLSLVKSVIYEFVSLQES 783

RESULT 8
 Q7XLE0 PRELIMINARY; PRT; 683 AA.
 AC Q7XLE0;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE OSJNBa0070C17.8 protein.
 GN Name=OSJNBa0070C17.8;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12447439;
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
 RA Liu Y., Hu X., Jia P., Zhang Q., Ying K., Yu S., Tang Y.,
 RA Weng Q., Zhang L., Lu Y., Mu J., Zhang L.S., Yu Z., Fan D.,
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
 RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
 RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
 RA Chen J., Kang H., Shao C., Sun X., Hu Q., Zhang X., Zhang W.,
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
 RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
 RA Han B.;
 RA "Sequence and analysis of rice chromosome 4.";
 RL Nature 420:316-320(2002).
 DR EMBL; AL731610; CAB05201.3; -.
 DR Gramine; O7XLE0; -.
 DR GO; GO:0016829; F-lyase activity; IEA.
 DR GO; GO:0008152; P-metabolism; IEA.
 DR InterPro; IPR005630; Terpene synthase C.
 DR InterPro; IPR001906; Terp synthase-like.
 DR Pfam; PF01397; Terpene synthase 1.
 DR Pfam; PF03936; Terpene synthase C; 1.
 SQ SEQUENCE 683 AA; 77861 MW; B18A2387456CAB84 CRC64;

Query Match 37.8%; Score 1556.5; DB 2; Length 683;
 Best Local Similarity 43.5%; Pred. No. 2.3e-83;
 Matches 323; Conservative 114; Mismatches 202; Indels 103; Gaps 12;

QY 48 RIQKQFNKNEIVSYSDTAWAVVPSNSPKSPCFPECLNWLNNQNDGSGWLVNHTN 107
 Db 19 RTDQQLTSLSLVDYDWTAWAVVPLGSRQHFQCFQVENVILNQDQDSWG----TRG 74
 QY 108 HNHPLKXDSLSTLACTIVALKRWNVGSDQINKGLSFIESMLAGATESQSPGFDIIPP 167
 Db 75 FGAVTRDVLSTLACVALKRWNVGQEHIRRGDLGFRNFSTAMDQIAAPVGFNITFP 134
 QY 168 GLLEYAKNLDINLLSKOTDFSLMHKELEKQR---CHSNEMDGYLAVISEGLNLYDWN 224
 Db 135 GMLSLGMDLEFFPQRTDVRDLHLHREIELEREAGDHSYGRKAYMAVTEGLNLEWD 194

QY 225 MVKTYQMKNQSVNPSATAAAFIHQNPGLNLYNSLIDKFGNAVPTVYPHDLFRLSM 284
 Db 195 EIMFQKNGSFNCPSTTAAITLVHNDKALQYLCNLL----- 233
 QY 285 VTIERLGHSHFRVEIKNVLDVTRCWVERDQIFMDVVVTCALAPRLIRINGEYVSP-- 342
 Db 234 -----VCSSWLERDEEIMLDITTCAMAFRLRMNGHYHVSVE 270
 QY 343 -DPLAEITN-----ELAKADYAALETYHSHLYQEDLSSGKQLKLSADFLKEIISTDS 396
 Db 271 LSPVAEASFRSLQYLNKSLIELYKASKWY----- 305
 QY 397 NRLSKLIHKEVENAKFPINTGLERINTRNIQLNVNDNTRILKTTYHSSNISNTDYLRL 456
 Db 306 -----ALKFPYVTLDRDHRKNIERFDKDSQMLKTEYLLPH-ANQDILAL 351
 QY 457 AVEDFYTCOSIYREELKGLERWVVENKLDQKPAROKTAYCFPSVAATLSSPESLARS 516
 Db 352 AVEDFSSQSISYQDELNLYLEWCWKDEKLDQKPAROKLTTCYLSAAATIPFPELSEARIA 411
 QY 517 WAKNGILTVDVDDFDIGGTIDELTNLIQCVKKNVVDVKDCGSEHVRILFLAKDAICW 576
 Db 412 WAKNGVLTIVDDFDLGGSGKELENLIALVEKND-CHQEEFYSEQVRIVFSIYTVNQ 470
 QY 577 IGDAEFKQARDVTSVHIQTWLELNMMLREAIWTEDAYVPTLNEYMENAYVSFALGPV 636
 Db 471 LGAKASALQGRDVTKLITWLCLMESWTEAEWQRTKVTMEETMANAVVSFALGPV 530
 QY 637 KPAIYVGPKLSEIIVESSEYHNLFKLMSTQGRLLNDIHSFKREFKGNALVALHLSNG 696
 Db 531 LPTLYFVGPKLQEDVVRDHEYNELFRLNSTCGRLNDSQGFERESLEGKLSVSL-LVHH 589
 QY 697 ESGKVE-EWVEREMMMIKRKLMLKLIPEENGSIIVPRACKDAFWNMCHVLPFFYAND 755
 Db 590 SGGISIDHAKWKAQKSIDTSRNLRLVLVGEQGA-VPRPCKOLFKNCKIVFMFISRTD 648
 QY 756 GFTG-NTILDVTKDIYNPLVL 776
 Db 649 GFSSPKEMVSANNAVVKPEPLK 670

RESULT 9
 Q6Z5J6 PRELIMINARY; PRT; 821 AA.
 AC Q6Z5J6;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative ent-kaurane synthase B.
 GN Name=P0689H05.20;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005114; BADI7270.1; -.
 DR InterPro; IPR005630; Terpene synthase C.
 DR InterPro; IPR008949; Terpene synthase.
 DR InterPro; IPR008930; Terpene synthase.
 DR InterPro; IPR001906; Terp synthase-like.
 DR Pfam; PF01397; Terpene synthase 1.
 DR Pfam; PF03936; Terpene synthase C; 1.
 SQ SEQUENCE 821 AA; 92376 MW; 52557D85FE875B11 CRC64;

Query Match 37.2%; Score 1530; DB 2; Length 821;
 Best Local Similarity 41.2%; Pred. No. 1.1e-81;
 Matches 332; Conservative 151; Mismatches 259; Indels 64; Gaps 16;


```
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative diterpene cyclase.
GS Name=OsDTC1;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14675427;
RA Cho E., Okada A., Kemoku H., Otomo K., Toyomasu T., Mitsuhashi W.,
RA Sassa T., Yajima A., Yabuta G., Mori K., Oikawa H., Toshihima H.,
RA Shibuya N., Nojiri H., Omori T., Nishiyama M., Yamane H.;
RT "Molecular cloning and characterization of a cDNA encoding ent-cassa-
RT 12,15-diene synthase, a putative diterpene synthase, from suspension-cultured rice cells treated with a chitin
RT enzyme, from suspension-cultured rice cells treated with a chitin
RT elicitor.";
RL Plant J. 37:1-8 (2004).
DR EMBL; AB089272; BAC56714.1; -.
DR Gramene; Q85282; -.
DR GO; GO:0016829; P:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005630; Terpene synth. C.
DR InterPro; IPR008949; Terpenoid synth.
DR InterPro; IPR008930; Terp. cyc. toroid.
DR InterPro; IPR001906; Terp. synth.-like.
DR Pfam; PF01397; Terpene synth. 1.
DR Pfam; PF03936; Terpene synth. C; 1.
SQ SEQUENCE 830 AA; 92091 MW; 1F23A5A43B703835 CRC64;

Query Match 37.2%; Score 1530; DB 2; Length 830;
Best Local Similarity 43.0%; Pred. No. 1.1e-81;
Matches 324; Conservative 122; Mismatches 265; Indels 42; Gaps 13;

QY 49 IQQFKNVLSVSSYDTAWVAVPSPNSPKSPCFPECLNWLNNQNDGSGVLNHTNH 108
DB 92 IRKQGLVELSPSPYDTAWVAVPLRGSSHNPSFPQCVDMILENQDGDGSGSIDGSI 151
QY 109 NHPLLKDSLSSTLACIALVKRNVDGQINKGLSFIESNLASATEKSQSPFGDIFPG 168
DB 152 N----KDVLSSTLACIALVKRNVDGQINKGLSFIESNLASATEKSQSPFGDIFPG 207
QY 169 LLEYAKNLDINLSKQDTPSMLHKELEOKR---CHSNEMDGYLAYSISGLNLYDNWM 225
DB 208 MLTLANGSGLEVPRQNDIDSLNHLREMKIQREAGNHSRGRKAYMAYLAEFGNLLWE 267
QY 226 VKKYQMKNGSVFNSPSATAAFINHQNPCLNLYNSLLDKFGNAVPTVYPHDLFIRLSM 285
DB 268 IMFFQRNGSLFNCPSSTAGALANYHDDKALQYLSLVNKFQGVVPTLYPLNYCQLSMV 327
QY 286 DTLERLGI SHHFRVEIKNVLDIETRCVVERDEQIFMDVVTTCALAFRLRLNGEVSDD 345
DB 328 DALENMGI SQYFASIEIKSILDMTYSYSLWLGKDEIMLDVTTCAVAFRLRWNGYD 387
QY 346 AETNELA-----LKDEYAALETYHAS-HILYQEDLSGKQILKS-----ADPLKE 390
DB 388 SHVAGSGFRDLSQGYLNDKRSVLEVKYTKSHISENDL-----ILDSIGSGSGSLKE 442
QY 391 IISTDSNRLSKLHKVENALKEPINTGLERINTRRNIOLYNDVNTIRLKT---YHSSN 447
DB 443 MLS--SNKGTGPGEETEPALKYFPYFTSLERLVRKNVILFDAGKSQMLKTCWPHDSQ 500
QY 448 ISNTDYRLAVEDFYTCQSTYREELKGLERWVENKLDQKFAKQKTAICYFVAATLSS 507
DB 501 ----DFLAVADDPCISQSNQNELNYLGSWKDNLQDLHFARQKITCYCLSGAATFR 556
QY 508 PELSDARLSWAKGILITVDDPFDIGGTTDELTLNLCQVKNVVDKCCCEHVLIF 567
DB 557 PEMGYARTSWARTAWLTAVIDDLFDVGGSLGQENLLALMEKEEPEGEDEYISDVKIVF 616
```

```
QY 568 LALKDAICWIGDEAFKQWQARDVTSHVIOQTWLEIMNSMLREAIWTRDAYVPTLNEYMENAY 627
DB 617 QALYNTVNEIGAKASALQGHDTVKLYDVWLHVRCMKVEAWQSRSHLPTFEYEWESGM 676
QY 628 VSFALGPVIVKPAIYFVGPKLSEIIVSESSYHMLFKLMSTQGRLLNDIHSHKREFFKGL- 686
DB 677 VSLGGQCTVMSALFLIGEXLPGIIVELBEYDELRLMGTGCRLLNDIRGIERESDGKMT 736
QY 687 NAVAL--HLSNGESGKVEEVEVMMMIKNRKMELMKLIFPENGSIIVPRACKDAFWNMC 744
DB 737 NGVSLLVHASGGSMS--VDEAKTEVMKRIDAGRRKLLSLVSVSEQEGPIPRPCKQLFWKMC 794
QY 745 HVLNPFYANDGFTG-NTILDTVKDIIYNPLVL 776
DB 795 KILHLEFYOTDGFSSPKEMVSAVDVINEPLQL 827

RESULT 12
Q6YV94 PRELIMINARY; PRT; 802 AA.
AC Q6YV94;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative ent-kaurene synthase.
GN Name=OSJNB000801.23;
OS Oryza sativa (japonica cultivar-group),
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki I., Matsumoto T., Katayose Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005835; BAD17672.1; -.
DR InterPro; IPR005630; Terpene synth. C.
DR InterPro; IPR008949; Terpenoid synth.
DR InterPro; IPR008930; Terp. cyc. toroid.
DR InterPro; IPR001906; Terp. synth.-like.
DR Pfam; PF01397; Terpene synth. 1.
DR Pfam; PF03936; Terpene synth. C; 1.
SQ SEQUENCE 802 AA; 89665 MW; 1D45CB1FC46E42AC CRC64;

Query Match 37.1%; Score 1525.5; DB 2; Length 802;
Best Local Similarity 43.0%; Pred. No. 1.9e-81;
Matches 321; Conservative 122; Mismatches 273; Indels 31; Gaps 11;

QY 49 IQQFKNVLSVSSYDTAWVAVPSPNSPKSPCFPECLNWLNNQNDGSGVLNHTNH 108
DB 65 IRKQGLVELSPSPYDTAWVAVPLRGSSHNPSFPQCVDMILENQDGDGSGSIDGSI 124
QY 109 NHPLLKDSLSSTLACIALVKRNVDGQINKGLSFIESNLASATEKSQSPFGDIFPG 168
DB 125 N----KDVLSSTLACIALVKRNVDGQINKGLSFIESNLASATEKSQSPFGDIFPG 180
QY 169 LLEYAKNLDINLSKQDTPSMLHKELEOKR---CHSNEMDGYLAYSISGLNLYDNWM 225
DB 181 MLTLANGSGLEVPRQNDIDSLNHLREMKIQREAGNHSRGRKAYMAYLAEFGNLLWE 240
QY 226 VKKYQMKNGSVFNSPSATAAFINHQNPCLNLYNSLLDKFGNAVPTVYPHDLFIRLSM 285
DB 241 IMFFQRNGSLFNCPSSTAGALANYHDDKALQYLSLVNKFQGVVPTLYPLNYCQLSMV 300
QY 286 DTLERLGI SHHFRVEIKNVLDIETRCVVERDEQIFMDVVTTCALAFRLRLNGEVSDD 345
DB 301 DALENMGI SQYFASIEIKSILDMTYSYSLWLGKDEIMLDVTTCAVAFRLRWNGYD 360
QY 346 AETNELA-----LKDEYAALETYHAS-HILYQEDLSGKQILKSADFLKEIISTDS 396
DB 361 SHVAGSGFRDLSQGYLNDKRSVLEVKYTKSHISENDLILDSIGSGSGSLKMLC--S 418
QY 397 NRLSKLHKVENALKEPINTGLERINTRRNIOLYNDVNTIRLKT---TYHSSNISNTDY 453
```

```

Db 419 NGKGTGREIEIFALKYFYSTLTLELVRHKNIVLPDAKGSQMLKTECMFVHDSQ----DF 474
QY 454 LRLAVDEFYTCOSIYREELKGLERWVENKLDQKFAROKTAYCYPSVAATLSSPESLSDA 513
Db 475 LALAVDFCISOSNYQNELNYLESWKDNLRLDQLHFAKQKITCYLSGAATFRPEMGYA 534
QY 514 RISWAKNGILITVDDFFDIDGTTIDELTNLIOCVEKNVVDKDCSEHVRILFLALAKDA 573
Db 535 RTSWARTAWLTAVIDDLDVVGLEQEQENLLALMEKWEPEGDEYSEYSDVKIVFOALYNT 594
QY 574 ICWIGDEAFKQARDVTSHVITQWLELMSMLREAIWTRDAYVPTLNEYMENAYVSFALG 633
Db 595 VNEIGAKASALQGHVDVTKYLVWLVHVRCKMVEAEWQSQHLPTFEETMESGMVSLGOG 654
QY 634 PIVKPAIYFVGPKLSEIEVESSEYHNLFKLMSTQGRLLNDIHSFKREFKGL-NAVAL- 691
Db 655 CTVMASALFLIGEKLPFGIVLEEYDFELFRLMGTQGRLLNDIRGIEREESDGRWNTGVSLL 714
QY 692 -HLSNGESGKVEEVVEEMMMIKNKRKELMKLIIFEENGSIIVPRACKDAFNMCHVLNFF 750
Db 715 VHASGGSMS--VDEAKTEVMKRIDASRRKLLSLVSEQEGPIPRPCKQLFWKCKILHLF 772
QY 751 YANDGFTG-NILDTVVDKDIINPLVL 776
Db 773 YYQTDGFSPPKEMVSAVDVAVINEPLQL 799

RESULT 13
BADI7672
AC BADI7672 PRELIMINARY; PRT; 802 AA.
DT 10-MAY-2004 (TREMELrel. 27, Created)
DT 10-MAY-2004 (TREMELrel. 27, Last sequence update)
DE Putative ent-kaurene synthase.
GN OSUNBA008E01.23.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza; Oryza sativa.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone:OSUNBA008E01."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005835; BADI7672.1; -.
SQ SEQUENCE 802 AA; 89665 MW; 1D45CB1FC46E42AC CRC64;

Query Match 37.1%; Score 1525.5; DB 2; Length 802;
Best Local Similarity 43.0%; Pred. NO. 1.9e-81;
Matches 321; Conservative 122; Mismatches 273; Indels 31; Gaps 11;

QY 49 IQKQFNVEISVSSYDTAWAVPSPKSPCEPCLNLLNQLNDGSLVNHTEH 108
Db 65 IRKQLQGVLESPSYDTAWAVPVRGSHNPSFPQCDVWILENQDQGSIDGSISTA 124
QY 109 NHPLLKSLSTLACTIVALKRWNVGEDQINKGLSFIESNLASATEKSPSPIGFDIIFPG 168
Db 125 N----KDVLSSTLACVLALNKNVNGREHRRGLSFIENFSIAMDQDAVPIGIGITFA 180
QY 169 LLEVAKNLDINLLSKQDFTSLMLHRELEQKR----CHSNEMDGYLAIYSEGLNLYDNW 225
Db 181 MLTLANGSGLEVPVRQNDIDSINHLREMKIQREAGNHSRGRKAYMAYLAEGFGLLEWDE 240
QY 226 VKKQMKNGSVNPSATAAFINHQNPCLNLYNLSLLDKGNVAVPTVYPHDLFIRLSMV 285
Db 241 IMVFQRKNGSLFNCPSSFTAGALANVHDDKALQYLSLVNKGFDGVVPTPLNIYCOLSMV 300
QY 286 DTIBELGISHFVRVEIKNVLDVTCWVERDEQIFMDVVTTCALAFRLIRINGEVSPDPL 345

```

```

Db 301 DALENWGISQYFASEIKSILDMTYSSWIGDKBEIMLDVTTCAWAFRLRNWGYDVSSDEL 360
QY 346 ABITNELA-----LKDEYAALETIHAS-HILYQEDLSSGKQILKSAADFLKEIISTDS 396
Db 361 SHVAGASGFRDLSQGVLDNRKSVLEVYKTSKHSISENDLILDSIGSWSSGLAKEMLC--S 418
QY 397 NELSILIKHEVENALKPPTINTGLERINTRNITQLVNVDNTRILKT---TYHSSNISNTDY 453
Db 419 NGKGTGREIEIFALKYFYSTLTLELVRHKNIVLPDAKGSQMLKTECMFVHDSQ----DF 474
QY 454 LRLAVDEFYTCOSIYREELKGLERWVENKLDQKFAROKTAYCYPSVAATLSSPESLSDA 513
Db 475 LALAVDFCISOSNYQNELNYLESWKDNLRLDQLHFAKQKITCYLSGAATFRPEMGYA 534
QY 514 RISWAKNGILITVDDFFDIDGTTIDELTNLIOCVEKNVVDKDCSEHVRILFLALAKDA 573
Db 535 RTSWARTAWLTAVIDDLDVVGLEQEQENLLALMEKWEPEGDEYSEYSDVKIVFOALYNT 594
QY 574 ICWIGDEAFKQARDVTSHVITQWLELMSMLREAIWTRDAYVPTLNEYMENAYVSFALG 633
Db 595 VNEIGAKASALQGHVDVTKYLVWLVHVRCKMVEAEWQSQHLPTFEETMESGMVSLGOG 654
QY 634 PIVKPAIYFVGPKLSEIEVESSEYHNLFKLMSTQGRLLNDIHSFKREFKGL-NAVAL- 691
Db 655 CTVMASALFLIGEKLPFGIVLEEYDFELFRLMGTQGRLLNDIRGIEREESDGRWNTGVSLL 714
QY 692 -HLSNGESGKVEEVVEEMMMIKNKRKELMKLIIFEENGSIIVPRACKDAFNMCHVLNFF 750
Db 715 VHASGGSMS--VDEAKTEVMKRIDASRRKLLSLVSEQEGPIPRPCKQLFWKCKILHLF 772
QY 751 YANDGFTG-NILDTVVDKDIINPLVL 776
Db 773 YYQTDGFSPPKEMVSAVDVAVINEPLQL 799

RESULT 14
QYXX63
ID QYXX63 PRELIMINARY; PRT; 770 AA.
AC QYXX63;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE OSUNBA0052P16.14 proteoin.
GN Name=OSUNBA0052P16.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12447439;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Wang Q., Zhang L., Lu Y., Wu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Zhu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Iv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL662936; CAD39717.1; -.
DR Gramene; Q7XX63; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005630; Terpene synth. C.
DR InterPro; IPR001906; Terp synth-like.
DR Pfam; PF01397; Terpene synth. 1.
DR Pfam; PF03936; Terpene synth. 2.

```

SQ SEQUENCE 770 AA; 86960 MW; 6CFAB6855D904EE1 CRC64;

Query Match 32.1%; Score 1312; DB 2; Length 770;
Best Local Similarity 37.7%; Pred. No. 2.7e-69;
Matches 288; Conservative 146; Mismatches 223; Indels 106; Gaps 16;

QY 32 GGCTNPTLIIDTKERIKQKFNVEISVSSYDVAWVMPGNSP--KSPCFPECLNWL 89
DB 89 GENTSQINM--GWEARIILHLENPEFLPSSDYDIANWVPLEGTHLQAPCEPECEVMI 145
QY 90 INNOLNDGSGWLNVNTHNNHLLKDSLSSTLACTIVALKRWVNGEDQINGLSFTIESNLA 149
DB 146 LQNHSGNSGWV---NEEDSASKDILLSTLACTIILEKWNVSGEIRRLGHFIKFNFS 201
QY 150 SATEKQSPGIDFIDFPGLELVAKNLDINLLSKQTDPSMLHKLREKQKCHSNE--M 206
DB 202 IVIDDQIAPIGNTFFPAMVNIAMKLEFPASEISIDQILHLRDMELKRLAGDSGLK 261
QY 207 DGLAYISEGL-GLNYDWNWVKYQMKNGSVFNSPATAAFAFTHNQPGCLNYLNSLLDK 265
DB 262 EAYFAIAGLESWVNDSEVMKFGKNGSLFNSPATAAALVHRYDDKALGYLYSVNKK 321
QY 266 FGNVPTVPHDLFIRLSWDTIERIGISHHFVEIKNVLDIETVRCVWERDEQIFMDVVT 325
DB 322 FGGEVPTVPLNIFSQLSMVDTLVNIIGISRHFSDDIKRIIDTKTYILWSQDEEWMLDLP 381
QY 326 CALAPRLIRINGEVPDPLAEITN-----ELAKDEYAALETYVHASHILYQEDLS 377
DB 382 CAMAFLLRMNGVGVSSDDLHVAEASTFNSVEGYLDLDTKSLLEYKASKV-----SLSE 437
QY 378 GKQILK-----SADFLKEIISTDSNRSLKIHEVENALKFPINTGLERINTERNIQLYN 432
DB 438 NEPILEKWCWSGLLKEKLCSDDIRGTPIL-REVEALKFPFYATILEPLDHKNENFD 496
QY 433 VDNTRILKTYHSNTSNTDYLRLAVEDFTYQSGIYREELKGLRWVVENKLDQLKPARQ 492
DB 497 ARAYQIKTKNMPCHV-NEDLLALAAEDFSCQSTYQNETQHLRW-----DQ----- 543
QY 493 KTAICYFSYAATLSSPELSDARTSWAKNGLLITVVDVDFDIDGTTIDELTNLIQCCKWV 552
DB 544 -----YHKV-----EFY----- 550
QY 553 DVDKDCSSEHVRILFLALKDAICWIGDEAFKQARDVTSHVIOTWLELMSMLREAIWTR 612
DB 551 -----SENVKAVFALYSTVNLGAMASAVQNROVTKYNVESWLDYLSLATDAEWQR 603
QY 613 DAVPTLNEWENAYVSFALGPVIVKPAIYFVGPKLSEIEVESSEYHNLFKLMSTQGLLN 672
DB 604 SKYVPTMEETKNSIVTFALGPITLIALYFMQNLWEDIVKNAEYDFELFRIMNTCGLQN 663
QY 673 DIHSEKPEKEGKINAVLHLSNGESGKVEEVEEMMMIKNKKELMKLIPEENGSI 732
DB 664 DIOSFERCKDGKINSVSLVDSKOVMSVEEAKENESISSCRRELLRLVWREDG-VI 722
QY 733 PRACKDAFNMCHLVNFPFVNDGFTG-TNLTDTVKDIIYNPL 774
DB 723 PKSCKEFNWLYKTSVHYFYQADGFSPKEMGMGAVNGVIEPL 765

RESULT 15

Q7XLD9 PRELIMINARY; PRT; 913 AA.

AC Q7XLD9

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE OSJNBa0070C17.9 protein.

GN Name=OSJNBa0070C17.9;

OS Oryza sativa (japonica cultivar-group).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI_TaxID=39947;

RP SEQUENCE FROM N.A.

RX PubMed=12447439;

RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J., Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y., Wang Q., Zhang L., Lu Y., Mu Y., Sun T., Lei H., Li T., Hu H., Guan D., Liu X., Lu T., Li C., Wu Y., Sun T., Chen L., Jin Z., Wang R., Yin H., Wu M., Zhang R., Zhou B., Chen Z., Chen L., Tu Y., Jia J., Zhang Y., Cai Z., Ren S., Lv G., Gu W., Zhu C., Sun Y., Hu Q., Zhang X., Zhang W., Chen J., Kang H., Chen X., Shao C., Sun Y., Chen S., Ni L., Zhu F., Chen W., Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W., RA Han B., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y., RT "Sequence and analysis of rice chromosome 4.";

RL Nature 420:316-320(2002).

DR EMBL; AL731610; CAE05202.3; --

DR Gramene; Q7XLD9; --

DR GO; GO:0016829; F-lyase activity; IEA.

DR GO; GO:0008152; Pimetalolism; IEA.

DR InterPro; IPR005630; Terpene synth_C.

DR InterPro; IPR001906; Terp synth-like.

DR Pfam; PF01397; Terpene synth_1.

DR Pfam; PF03936; Terpene synth_C; 1.

SQ SEQUENCE 913 AA; 99830 MW; 3FF5D613C92C6C7F CRC64;

Query Match 31.9%; Score 1312; DB 2; Length 913;
Best Local Similarity 38.4%; Pred. No. 9e-69;
Matches 289; Conservative 142; Mismatches 232; Indels 90; Gaps 20;

QY 70 MTPS-PNSPKSPCFPECLNWLINQLNDGWS-GLVNHNNHNP--LLKDSLSSTLACI 124
DB 1 MVDALRGGGGPRFPQCVAMIQNRQDGSVHAAAHQOLGSSPRIVTERDLSSTLACV 60
QY 125 VALKRWVGEQDQINKLSFTESNLASATEKSPSPGFDII-PPGLELVAKNLDINLLS- 182
DB 61 LALARWDAGEHVRRLGQIFGRNNSVAMDQTPAPASGVSVAAMLRWAMGLEVPVAV 120
QY 183 -----KQTFSLMLHKLRELEQKCHSNEMDGYLAYISEGLNLYDNWVKYQMKNGSVF 237
DB 121 SQADVDRDRDAGVI-----CHGRTE-YTAYVSEGLGNTQNNWVNMKFORKNGSLF 169
QY 238 NSPSATAAFATFNQNGCLNLYNSLIDKFGNAVPTVYPHDLFIRLSWDTIERIGISHHF 297
DB 170 NSPTTAAALVHNVDAKALQYLDMLDKFGSAVPAAPANIQOLYQVNDVLEKMGIRHF 229
QY 298 RVKIKVLDIETVRCVWERDEQIFMDVVTALAPRLIRINGEVPDPLAEIT-NE 351
DB 230 VGIKSIILDMTYSQWQORDEIVLDMQTCGMAFRLMNGYDVSSELSHSESPSSPHNS 289
QY 352 LA--LKDEYAALETYVHASHILYQEDLSGKQILKSDAFLKEIISTDSNRSLKIHEVEN 409
DB 290 LQGYLNDTRSLLELHASKV-----STAEKEVEY 318
QY 410 ALKFPINTGLERINTERNIQLYNVDNTRILKTYHSNINSDTYLRLAVEDFTYQSGIYR 469
DB 319 ALEFFPFTILDRLDHKNIEHFDITSQMLETAYLPCH-SNEETMALGVDFSSSQIFQ 377
QY 470 EELKGLRWVVENKLDQLKPARQKAYCFVSAATLSSPELSDARTSWAKNGLLITVVDV 529
DB 378 EELQOLNSWKSESLDQLQFARQKLDYFSAATITFPELSDVRLWAKNGVLTIVDD 437
QY 530 FFDIGGTTIDELTNLIQCCKWVNDVKDCOCCSEHVRILFLALKDAICWIGDEAFKQARDV 589
DB 438 FFDVGSKEELENLVALVEKWDKNDKTEYYSEQVEIVFSAITYSTNLGSMASVQGRDV 497
QY 590 TSHVIQTWLELMSMLREAITW---RDAYVPTLNEYME-NAYVVSFALGPI-----VK 637
DB 498 TKHLE-----IVSIISDILWNVFFSVRTDKNIVASVNLCAVARIAEYVDDRCRVEAE 552
QY 638 PAIFYVGPKLSEIEVESSEYHNLFKLMSTQGLNDIHSFKPEKGLKINAVLHLSNGE 697
DB 553 PHCILLGPKMPDSVIRSQCESELFRLMKSGRLLNDVQSVYEREGSQGLKNSVSL-LALHS 611

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:05:06 ; Search time 51.1457 Seconds
(without alignments)
1474.884 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSLCIASPLTKSNRPA.....TVKDIYPLVNVNEBQR 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------|----------------------|
| 1 | 2023.5 | 49.2 | 789 | T09672 | ent-kaurene synthase |
| 2 | 2021 | 49.1 | 785 | G96825 | hypothetical prote |
| 3 | 2021 | 49.1 | 785 | T52059 | ent-kaurene synthase |
| 4 | 813.5 | 19.8 | 870 | A96637 | hypothetical prote |
| 5 | 725.5 | 17.6 | 801 | T06783 | ent-kaurene synthase |
| 6 | 709 | 17.2 | 802 | T02959 | ent-kaurene synthase |
| 7 | 657.5 | 16.0 | 823 | T02959 | kaurene synthase A |
| 8 | 446.5 | 10.9 | 591 | H84633 | probable limonene |
| 9 | 431 | 10.5 | 600 | T71434 | probable limonene |
| 10 | 426 | 10.4 | 559 | T08174 | sesquiterpene cycl |
| 11 | 413 | 10.0 | 599 | A48863 | limonene cyclase - |
| 12 | 406 | 9.9 | 1024 | T71434 | probable limonene |
| 13 | 399 | 9.7 | 550 | T03714 | 5-epi-aristolochene |
| 14 | 398.5 | 9.7 | 520 | A56118 | vetispiradiene syn |
| 15 | 373 | 9.1 | 554 | S68366 | (+)-delta-cadinene |
| 16 | 367 | 8.9 | 554 | S68385 | germacrene C synth |
| 17 | 357.5 | 8.7 | 548 | T06285 | hypothetical prote |
| 18 | 354 | 8.6 | 548 | T06266 | hypothetical prote |
| 19 | 354 | 8.6 | 632 | E96723 | probable terpene s |
| 20 | 326 | 7.9 | 608 | T06285 | hypothetical prote |
| 21 | 321 | 7.8 | 598 | H86460 | hypothetical prote |
| 22 | 290 | 7.1 | 535 | T06285 | hypothetical prote |
| 23 | 288 | 7.0 | 530 | T06285 | hypothetical prote |
| 24 | 278 | 6.8 | 598 | T00509 | probable vetispira |
| 25 | 270.5 | 6.6 | 471 | T06287 | hypothetical prote |
| 26 | 267 | 6.5 | 421 | C96642 | hypothetical prote |
| 27 | 260 | 6.3 | 612 | T05331 | hypothetical prote |
| 28 | 253 | 6.2 | 573 | T05328 | hypothetical prote |
| 29 | 251 | 6.1 | 582 | C71424 | hypothetical prote |

| | | | | | | |
|----|-------|-----|------|---|--------|----------------------|
| 30 | 249 | 6.1 | 350 | 2 | C56118 | vetispiradiene syn |
| 31 | 244.5 | 5.9 | 598 | 2 | T05329 | hypothetical prote |
| 32 | 232.5 | 5.7 | 501 | 2 | H70972 | probable cyclase - |
| 33 | 230.5 | 5.6 | 598 | 2 | F96684 | probable terpene s |
| 34 | 228 | 5.5 | 946 | 2 | T00024 | ent-kaurene synthase |
| 35 | 207.5 | 5.0 | 300 | 2 | B56118 | vetispiradiene syn |
| 36 | 203 | 4.9 | 383 | 2 | D71424 | hypothetical prote |
| 37 | 199.5 | 4.9 | 952 | 2 | JC7227 | hypothetical prote |
| 38 | 157 | 3.8 | 2010 | 2 | B71616 | phosphatase (acid |
| 39 | 156 | 3.8 | 1538 | 2 | T29095 | cardiac muscle fac |
| 40 | 155 | 3.8 | 3351 | 2 | T13812 | lipophorin - fruit |
| 41 | 151 | 3.7 | 753 | 2 | S22802 | transposase (clone |
| 42 | 149.5 | 3.6 | 2335 | 2 | T40186 | probable phosphat |
| 43 | 148 | 3.6 | 3394 | 2 | T18501 | hypothetical prote |
| 44 | 147 | 3.6 | 2166 | 2 | G70163 | hypothetical prote |
| 45 | 144 | 3.5 | 971 | 2 | A70179 | exodeoxyribonuclea |

ALIGNMENTS

RESULT 1

T09672
ent-kaurene synthase B (EC 2.5.1.-) - winter squash

C:Species: Cucurbita maxima (winter squash)

C:Date: 16-Jul-1993 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T09672

R:Amaguchi, S.; Saito, T.; Abe, H.; Yamane, H.; Murofushi, N.; Kamiya, Y.

Plant J. 10, 203-213, 1996

A:Title: Molecular cloning and characterization of a cDNA encoding the gibberellin biosyn

A:Reference number: Z16814; MUID:96367664; PMID:8771778

A:Accession: T09672

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-789 <YAM>

A:Cross-references: UNIPROT:Q39548; EMBL:U43904; NID:G1431869; PIDN:AAB39482.1; PID:G143.

A:Experimental source: immature seeds

C:Function:

A:Description: catalyzes the conversion of copalyl diphosphate to ent-kaurene

A:Pathway: gibberellin biosynthesis

A>Note: terpene cyclase

C:Superfamily: terpene synthase

C:Keywords: transferase

Query Match 49.2%; Score 2023.5; DB 2; Length 789;
Best Local Similarity 50.0%; Pred. No. 1.2e-112;
Matches 393; Conservative 150; Mismatches 212; Indels 31; Gaps 8;

| | | |
|----|-----|---|
| QY | 16 | NRPAALSAIHTASTSHGGQTNPTNLIIDT-----TKERIQKQKQKVEISVSVD 64 |
| DB | 4 | SRPTGVARFAASSSSSSSSASLFPQGVDDTTTKTGALHFEETKERIKLFDKVELSVSAYD 63 |
| QY | 65 | TAWVAVPSPNSPKSPCFPECLNWLNNQNDGSGWLVNHNHNPILL-KDLSLSSTLAC 123 |
| DB | 64 | TAWVAVPSPNSLNQPLFPPECINWVLSQADSGWLL-----HNDQLLMKANLLSTLAC 118 |
| QY | 124 | IVALKRWVVEDQINKGLSIESNLASATKQSPGTGDIIFPGLLLEYAKNLDINLLSK 183 |
| DB | 119 | VLTKRWVIGHDHMSKALDFIKSNIASATDENQSPVGFDIIFPGMIEYAKNLDINLPLA 178 |
| QY | 184 | QTDPSLMHLKRELQKCHSNEMDG--YLAVISGLNLYDMNWKYQKNGSVFNSPS 241 |
| DB | 179 | FTNVDAIVRKKELELRCSRNSGKKAYLAYSEGIQKQNDWMVMQYKQNGSLFNSPS 238 |
| QY | 242 | ATAAAFINHQNGPCINYLNSLDKFGNAVPTVYPHDLFIRLSMVDTTIRLGISHFRVEI 301 |
| DB | 239 | TTRAAFMHRNDGCFYLRSLQKFDGSVFTIPLDIYARLHMVDSLQKFGIARHFRKEI 298 |
| QY | 302 | KNVLDETYRCWVRDEQIFMDVVTCAALAFLLRNGYEVSPPDLAEITNEL-----AL 354 |
| DB | 299 | RSVLDETYRCWQGEENIFLDASTCAMAFLKRVGVDVSSDQLTFQSEDIFFNCLGYYL 358 |
| QY | 355 | KDEVALETYHASHILYQEDLSGKQILK-SADFLKEIISTD---SNRLSKLIHKEVENA 410 |

Db 359 KDFASLELYKASQIITHPDESVLNINSWTSFLKHGLSSDSWSDRTSDSVKQEAUNA 418
 Qy 411 LKFPINTGLERINTERNITQLNVNDTRILKTTTHSSNISTNTDYLRLAIVEDFTCOSIYRE 470
 Db 419 LEFFYNATLERLISKRAMESYSGDIVRISKSPYACLNFGHQDFLELAIVEDFNTLQRIHLK 478
 Qy 471 ELKGLERWVENKLDOLKPAROKTAYCYPVSAATLSSPELSDARISWAKNGILTTVVDDP 530
 Db 479 ELEELQWVENKLDOLKPAROKTAYCYPVSAATLSSPELSDARISWAKNGILTTVVDDP 538
 Qy 531 PDIGTIDELTNLICOVKEWVDDKCCSEHVRILFLALDKAI CWIGDEAFKQOARDVT 590
 Db 539 YDGGSEBELDLNILELVEKWDPDGEVGYCKDVEIVFLAHSTVEIGERRALVQGRSV 598
 Qy 591 SHVITQWLELMSMLREAIWTRDAYVPTLNEYMENAYVSFALGPVIVKPAIYVGPKLSEE 650
 Db 599 RNVIDGWLALLKVRKEAEWSTNKVPSMGMEYQAHVSFALGPVILPMLFFVGPKLSEE 658
 Qy 651 IVSESEYHNLFKLMSTOGLINDIHSFKREFKEGKLNALVALHLSNGESKVEEVEVWMM 710
 Db 659 MIGSEYQYLYKMLSTAGLKNDIRSYDRECKEGLNLSLWMDGGGNVTKEEALEAK 718
 Qy 711 WMKKNRKLMLKLIPEENGSIYPRACKDAFWNMCHVLPNFYANDDGTGNTILDITVKDII 770
 Db 719 GDFERARELLGLVLOEN-TTIPRACKDLFWKLSIVNLFYMEDDGTNSRLMNTVKAMF 777
 Qy 771 YNPLVL 776
 Db 778 EQPMDL 783

RESULT 2
 G96825
 hypothetical protein T9K14.12 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C;Accession: G96825
 R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.P.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; MUID:21016719; PMID:11130712
 A;Accession: G96825
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-785 <STO>
 A;Cross-references: UNIPROT:Q9SAK2; GB:AB005173; NID:g4835764; PIDN:AAD30231.1; GSPDB:GN
 C;Genetics:
 A;Gene: T9K14.12
 A;Map position: 1
 C;Superfamily: terpene synthase

Query Match 49.1%; Score 2021; DB 2; Length 785;
 Best Local Similarity 52.6%; Pred. No. 1.7e-112;
 Matches 402; Conservative 130; Mismatches 208; Indels 24; Gaps 9;

Qy 34 QTNPTNLIIDTTERIKQKFNVEISVSYDTAWAVPSPKSPCFPECLNWLINQ 93
 Db 27 QTRANNVSFEQTEKIRKMLEKVELSVAYSAYDTSNVAVPSPSQNAFLFPQCVKWLNDQ 86

Qy 94 LNDGSGVGNHNNHNNHPLKDSLSSTLACTIVALKRWNVGDDQINKGLSFIESLASATE 153
 Db 87 HEDGSGVGNHNNHNNHPLKDSLSSTLACTIVALKRWNVGDDQINKGLSFIESLASATE 143
 Qy 154 KSQSPGIGFDIIFPGKIKYARDLNTIPLGSEVDD--MIRKRDLDLK-CDSEKFSKGRE 200

Db 144 ETIQKPTGFDIIFPGKIKYARDLNTIPLGSEVDD--MIRKRDLDLK-CDSEKFSKGRE 200
 Qy 208 GYLAVISGLGNLNDVNMVKKYQKNGSVFNSPSATAAATFQNGDGLCLYLCSSLLQKFE 267
 Db 201 AYLAVLVEGTNRLKDWDLIVIKIQRKNGSLFDSPTATAAFTQFGNDGCLYLCSSLLQKFE 260
 Qy 268 NAVPTVYVPHDIFRLSMVDITIERLIGISHHFRVEIKRNLDETVCYRWCVRDEQIFMDVVVTA 327
 Db 261 AAIVSVYFPDQVARIISIVTLBSLIGDRDKTEIKSILDETVCYRWCVRDEQIFMDVVVTA 320
 Qy 328 LAFRLRLINGEVSPDPLAEITNELA-----LKDEYAALETTHASHILYQEDLSSEK 379
 Db 321 LAFRLRLAHGVDSVDPKPFABESGSDTLEGYKNTFSVLELKPAAQ-SYPHESALKK 379
 Qy 380 QILKADFLKEITIST--DSNRLSLIHKHEVENALFPINTGLERINTERNIOLYN-VNDT 436
 Db 380 QCCWTKQYLEMELSSWVTSVDRDKVLEVEDALAFPSYASLERSDHRKILNGSAVENT 439
 Qy 437 RILKTYTHSSNISTNTDYLRLAIVEDFTYQSIYREELKGLERWVENKLDOLKPAROKTAY 496
 Db 440 RYTKTSYRLHNICTSDILKLAVDDEFNFCOSIHREMERLDRWVENRLOELKFAKOLAY 499
 Qy 497 CYFVSAATLSSPELSDARISWAKNGILTTVVDDPDI GGTIDELTNLICOVKEWVDDVK 556
 Db 500 CYFSAATLSSPELSDARISWAKNGILTTVVDDPDI GGTIDELTNLICOVKEWVDDVK 559
 Qy 557 DCCSEHVRILFLALDKAI CWIGDEAFKQOARDVTSHVITQWLELMSMLREAIWTRDAYV 616
 Db 560 RYSSEHVEIIFSVLRDITLETGDKAFTYQGRNVTHHIVKILDLKLSMLREAEWSSDKST 619
 Qy 617 PTLINEYMENAYVSFALGPVIVKPAIYVGPKLSEEVSESEYHNLFKLMSTOGLINDIHS 676
 Db 620 PSLEDYMENAYVSFALGPVIVKPAIYVGPKLSEEVSESEYHNLFKLMSTOGLINDIHS 679
 Qy 677 FRREFKEGKLNALVALHLSNGESKVEEVEVWMMKKNRKLMLKLIPEENGSIYPRAC 736
 Db 680 FKRESAEGKLNALVALHLSNGESKVEEVEVWMMKKNRKLMLKLIPEENGSIYPRAC 739
 Qy 737 KDAFWNMCHVLPNFYANDDGTGNTILDITVKDIIYNPLVLNEN 780
 Db 740 KEAFLKMSVNLNLFYRKDDGFTSNDLSLVKSVIYFVSLQES 783

RESULT 3
 T52059
 ent-kaurene synthase (EC 2.5.1.-) [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004
 C;Accession: T52059
 R;Yamaguchi, S.; Sun, T.; Kawaide, H.; Kamiya, Y.
 Plant Physiol. 116, 1271-1278, 1998
 A;Title: The GA2 locus of Arabidopsis thaliana encodes ent-kaurene synthase of gibberellin
 A;Reference number: Z25926; MUID:98205064; PMID:9536043
 A;Accession: T52059
 A;Status: preliminary;
 A;Molecule type: mRNA
 A;Residues: 1-785 <YAM>
 A;Cross-references: UNIPROT:O64952; EMBL:AF034774; PIDN:AAC39443.1
 C;Genetics:
 A;Gene: GA2
 C;Superfamily: terpene synthase
 C;Keywords: transferase

Query Match 49.1%; Score 2021; DB 2; Length 785;
 Best Local Similarity 52.6%; Pred. No. 1.7e-112;
 Matches 402; Conservative 130; Mismatches 208; Indels 24; Gaps 9;

Qy 34 QTNPTNLIIDTTERIKQKFNVEISVSYDTAWAVPSPKSPCFPECLNWLINQ 93
 Db 27 QTRANNVSFEQTEKIRKMLEKVELSVAYSAYDTSNVAVPSPSQNAFLFPQCVKWLNDQ 86

Qy 94 LNDGSGVGNHNNHNNHPLKDSLSSTLACTIVALKRWNVGDDQINKGLSFIESLASATE 153

Db 87 HEDGSGLDNDHDK---QSLKKOVLSSLASILAKKXGIGERQINQKGLQFIELNSALVTD 143
QY 154 KQSPSPGFDIIIFPGLLEYAK--NLQINILSKQTDPSLMHKKRELEBKCRKCHSNEM-----D 207
Db 144 ETIQKTPGFDIIIFPGMIKYARDNLNIPLGSEVDD--MIRKQDLIK-CDSEKFKSGRE 200
QY 208 GYLAYISEGLGNIYDMWVKYQKNGSVNPSATAAFINHPONPGCLNLSLADKFG 267
Db 201 AYLAYLEGTRNLKQWDLIVKYQKNGSLFSDPATAAFTQPGNDGCLRYLCLLQKFE 260
QY 268 NAVPTVYPHDLFRLSMDVTEIRLGSHHFRVEIKNVLDYTRCVRDEQIFMDVVTC 327
Db 261 AAVPSVYPDQVARUSIIVTLSEGLDRDFKTEIKSILDTYRVLGRDEEICLDLATCA 320
QY 328 LAFRLRLNGYEVSPDPLAEITNELA-----LKDEYAALETYHSHILYQEDLSSCK 379
Db 321 LAFRLLLAHGYDVSDPLKPAEESGFSITLGYVKNTPSVLELFKAAQ-SYPHESALKK 379
QY 380 QILKSADEFLKEIIST--DSNRLSKLIHKEVENALKFPINTGLERINTRRNIOLYN-VDNT 436
Db 380 QCCWTKQYLEMELSSWVKTSVRDKVLKKEVEDALAPPSVASLERSDHRKILNGSAVENT 439
QY 437 RILKTYHSSNTSDYLRLAYEDEVTCOSIYREELKGLRWVVENKLDOLKFAKQKAY 496
Db 440 RYTKSYRLNICTSDIILKLVDDNFNFCOSIHREMERLDRWIVENRLQELKFAKQKAY 499
QY 497 CYFSVAATLSSPELSDARISWAKNGILTTVDFFDGTIGTIDELTNLIQCEKKNVVDVK 556
Db 500 CYFSGAATLSPELSDARISWAKGGVITVDFFDVGSGKELENLHILVEKWDOLGVP 559
QY 557 DCCSHVRLFLAKDAICWIGDEAFKQOARDVTSHTVITQWLELNMNLSRLREAITRDAYV 616
Db 560 EYSSEHVEIIFSVLDDTILETGDKAFTYQGRNVTHIVKWLDLKSLMLRAEWSDDKST 619
QY 617 PTLNEMNAYVSPALGPVPAIYFVGPKLSEEVSEYHNLKLMSTQGRLLNDIHS 676
Db 620 PSLDYMENAYISPALGPVPAIYFVATYLIIGPEKTVDSHQVQKLYKLVSTWGLRLDIQG 679
QY 677 FKRFKQKLNALVHLNGESGKVEEVEEMMMKKNKRLMKLIFEENGSIIVPRAC 736
Db 680 FKRESAEKLNALVHLNGHEDNRSEKVIIESMKGLAEKREELHKLVLKESGVVPRC 739
QY 737 KDAFWNMCHVNLFFVANDGFTGNTILDTVKDIIYNPLVLNEN 780
Db 740 KEAFLKMSKVLNLFYRKDDGFTSNDLMSLVKSVIYEPVSLQES 783

RESULT 4

A96637
hypothetical protein Flp17.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A96637
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-870 <STO>
A;Cross-references: UNIPROT:O22733; GB:AEO05173; NID:G2443889; PIDN:AAB71482.1; GSPDB:GN

C;Genetics:

A;Gene: Flp17.15

A;Map position: 1

C;Superfamily: terpene synthase

Query Match 19.8%; Score 813.5; DB 2; Length 870;
Best Local Similarity 28.0%; Pred. No. 1.5e-40;
Matches 226; Conservative 163; Mismatches 322; Indels 97; Gaps 23;
QY 29 TSHGGGTNTNLIIDTTKERIQKQKNVE---ISVSSYDITAWVAVMPS-----PNSPKSP 80
Db 3 SYVGSSNDLHAFVNEIKGEIQLSNINLPYFVSPSAYDTAWLSMIEEDINVDDELKP 62
QY 81 CPPECLNWLINNO-LNDGSGWGLVNHHTNHNPLKDS-----LSSITLACIALVK 128
Db 63 MFQGCCLDMTCNQNAREGFW--MNST---SYITVADGRDEDEKOKCILTSYLCVVALQ 117
QY 129 RNWVGEDQINGLSPFESN---LASATEKSPSPIGFDIIIPFGLLEYAKNLDINLSKO 184
Db 118 KWNIGCFHLHKYTIERTNTEMIIGKYNIEGSPYRFAKFTGTILELAQKGLHFFVSS 177
QY 185 TDFSL---MLHRE---LEQKCHSNEMDGYLAYISEGLNLYDMN---MVKYQMKNG 234
Db 178 RCIEMLKGMFYQORBIQREKLVHDCNVKPLLAYLEVLPKLYVTNQEDIIIVKSLDSMDG 237
QY 235 SVFNSPSATAAFINHPONPGCLNLSLDFKGNAPVTYP-HDLFIRLSWVDVTIERLGI 293
Db 238 SLFQSPSATASAFMLTRNTKCLAYLQNLVQKCPGVPQKYPINEDLIKLSWNLIESGL 297
QY 294 SHHFRVEIKNVLDYTRCVRDE-----EQIFMDVVTICALAFRLRLNGYEVSPD 343
Db 298 GEFQIGIEHLVLEQVYSRYEKEDEKFMPSYLAQLQHKD---SLAFRLMRHGRDVSFR 353
QY 344 PLAEITNELALKDE-----YAALETYHSHILY--QEDLSSGKQ-----ILKSAFPL 388
Db 354 SFCWFLNDQETRNHLERNIDSLVLVSYRATDLMFPGEHDLQAREYTRNLLEKRSI 413
QY 389 KEIISTDSNRLSKLIHKEVENALKFPINTGLERINTRRNIOLYNVDNTRILKTTVHSSNI 448
Db 414 KE--KMWPIYLNKIMHE-----LSTFWIARLKHLDHRWIEDKSNVLSMEKASFLRLHS 466
QY 449 SNTDYL-PLAVEDFTQOSIYREELKGLRWVVENKLDOLKFAKQKAYCYFSVAATLSS 507
Db 467 SYSOKLTHLAARNFPOQAKYCRELEL-----TIEKTYCYFATVTSLPY 512
QY 508 PELSDARISWAKNGILTTVDFFDGTIGTIDELTNLIQCEKKNVVDVKCCSEHVRILF 567
Db 513 EYALFKGLAKTALITIAIDFFDEKGSFNDLEGLTAVLRWEGELK-----SYGNIIF 568
QY 568 LALKDAICWIGDEAFKQOARDVTSHTVITQWLELNMNLSRLREAITRDAYVPTLNEMNAY 627
Db 569 RALDDIVRETANTCKTHKTDIIIVHLRNIGETFSWLRREAEWSKKGHTSSMDEYIRNGM 628
QY 628 VSPALGPVPAIYFVGPKLSEEVSEYHNLKLMSTQGRLLNDIHSFKREPKFKGLIN 687
Db 629 ISIAAHTIALISICLMFPCFPFNKLPKGNVDSITTLMLIIPRLNLDLQSYQKEQOQGN 688
QY 688 AVALHLSNGESGKVEEVEEMMMKKNKRLMKLIFEENGSIIVPRACKDAFWNMCHVL 747
Db 689 SVLLHMKNHGLEIEDSIA-HIEKIDSKREFLEHVLVDGLSDLPKPKCEIHMSSCKVF 747
QY 748 NFFVANDGFTGNT-ILDTVKDIIYNPL 774
Db 748 EMFENKKNYDSNTMLHDIKKALYDPI 775

RESULT 5

T06783

ent-kaurene synthase A (EC 2.5.1.-) - garden pea

C;Species: Pisum sativum (garden pea)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C;Accession: T06783

R;Alt-Ali, T.; Swain, S.M.; Reid, J.B.; Sun, T.P.; Kamiya, Y.

Plant J. 11, 443-454, 1997

A;Title: The LS locus of pea encodes the gibberellin biosynthesis enzyme ent-kaurene synt

A;Reference number: Z15810; MUID:97260958; PMID:9107034

A;Accession: T06783

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-801 <Alt>
A:Cross-references: UNIPROT:004408; EMBL:U63652; NID:G2160543; PIDN:AAB58822.1; PID:G2160543
A:Experimental source: cultivar Torsdag
C:Genetics: LS
A:Gene: LS
C:Function:
A:Description: converts the geranylgeranyl diphosphate into copalyl diphosphate
A:Pathway: the first step of the gibberellin biosynthesis
C:Superfamily: terpene synthase
C:Keywords: transferase

Query Match 17.6%; Score 725.5; DB 2; Length 801;
Best Local Similarity 27.0%; Pred. No. 2.3e-35;
Matches 224; Conservative 145; Mismatches 281; Indels 181; Gaps 32;

QY 14 KGNRPAALSAIHTASTSHGGQTNPTNL-----IIDTTKER-----IQ 50
DB 39 KEKRCFAISKSRTOEYEGVFQINVATLKLSEINVEDVIVIDBEEQDIRVGLVNIKISIL 98
QY 51 KQPKNVEISVSYDTAWAMVSPSPKSPCTPECLNWLINQINDGSLVNHHTHHH 110
DB 99 SSLEDEGEITISAYDTAWVALVEDVNAISTPQFPSSLEWIAKNQLOQSGWSDRLFSAH-- 156
QY 111 PLIKDSLSSTIACTIVALKWNVCEQDQINKGLSFIESNLAS-ATEKSQSPPIGDIIFPGI 169
DB 157 ----DRIINTLACVIALRSWNHMSKCDKGMPIFRNLSKLENEHEHPIGFEVAFPSL 212
QY 170 LEYAKNLD-----INLASKQTFSLMLHKLRELEQKCHSNEMDGYLAIVISGLN 219
DB 213 LEGARGIKLPCNDSPILKNIPEKDEKLTIPKIMHK-----VPTLLHSLGMSG 266
QY 220 LYDNNMVKYQKMGKSVFNSPSATAAFINQHPGCLNVLNLLDKEGNAVPTVYDHLF 279
DB 267 L-DWKQLLKLQSDGSLFSPSSTAFALMOTKGNCLKYNVNVVKKFNGGVPNVYVDF 325
QY 280 IRLSMVDTTIRLGISSHFRVEIKNVLDITYRCWE-----RDEQIFMDVVTCALAFRL 333
DB 326 EHTVVDRLERLGLSRFRHIEIKDMVYVKIWKSEKGI CWARNV- QDIDDTAMAFRL 384
QY 334 RINGEYVSPDLAEITNELAKDEY-----AALETYHSHLYQEDLSGKQI 381
DB 385 RLHGHQVS-----AHVFKHFERNGEFCFAGCQTAQVGMVNLFRASQVLF-----PGEKI 435
QY 382 LK-----SADPLKEIISTDNRL-----SKLHKVEVENALKFPINTGLERINTRNIQL 430
DB 436 LEHAKHFSAKVKE--KREANELIDKWIIMKNLPEEVGYALDMPWYANLDRITRFYIDQ 493
QY 431 YNVD-NTRILKTTVHSSNITDYLRLAVDFYTCQSIYREBELKGLRWVVENKLDQLKF 489
DB 494 YGAEVWIGKTLRYMAYVNNNNLELAKLDYNNCOAHLIENWVIGTWLESGLGFFGL 553
QY 490 ARQKATCYFSAATLSSPELSDARISWAKNGILTTVVDDFDIGGPIDELTNLIQCVEK 549
DB 554 SKRDLLAYFLATGSIPEPERSHERLAWAK-----TTA-----LLETIKCYVR 596
QY 550 WNVVDVKDCC--SEHVRILFLAL-----KDAICWIGDEAFKQARDVTSH 592
DB 597 -NEDLRKDFAKFNDHDVDRDYSTARMKRNKTHSELVESLFAFIGE--ISWDVRLSYGH 653
QY 593 VI-----QTWLELMNSMLRE--AIWTRDAVPTLNEYMENAVVVSFALGPVVKPAIFYVG 644
DB 654 EIGYDMEQCKWKLSSWQSGDKCEGEAEILLIQIN-LCSNHWIS-----EG 699
QY 645 PKLSEEVSEVHNLPKLMSTQGRLLNDI-HSPKRFKQKGNALVAL--HLSNGRSGKV 701
DB 700 PSM-----QSTIQELL-----QLTNSICHKLSYQKQKELKGISQENITNSE---- 742
QY 702 BEEVVEENMMKRNKRLMKLIFEENGSIYPRACKDAFNMWCHVLNPFYA 752
DB 743 -----VESKMQELVQMVPQKCFNDIDFNVKNTFTFIATK--SPFYA 780

RESULT 6

D85035

ent-kaurene synthetase A-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: D85035
R:anonymouse, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: D85035

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-802 <STO>

A:Cross-references: UNIPROT:Q38802; GB:NC_001268; NID:G7269763; PIDN:CAB77763.1; GSPDB:G7269763

C:Genetics: A8502780

A:Gene: A8502780

A:Map position: 4

C:Superfamily: terpene synthase

Query Match 17.2%; Score 709; DB 2; Length 802;
Best Local Similarity 32.6%; Pred. No. 2.2e-34;
Matches 174; Conservative 92; Mismatches 216; Indels 52; Gaps 13;

QY 30 SHGGQTNPTNLIIIDTTKERIQKQKXNVEISVSYDTAWAMVSPSPKSPCTPECLNWL 89
DB 86 SVGSNSNAFKAQVSVK-TILRNLTGDEITISAYDTAWVALIDA--GDKTPAPPSAVKWI 142
QY 90 INNQLDGSGLVNHHTHHHPLIKDSLSSTIACTIVALKWNVCEQDQINKGLSFIESNL 149
DB 143 AENQLSDGSGWDAYLFSYH-----DRLNTLACVVALRSWNLFPHQCKNGITFFRENIG 196
QY 150 S-ATEKSQSPPIGDIIFPCGLLEYAKNLDINLLSKOTDFSLMLHKLRELEQKCHS--NE 205
DB 197 KLEDENDHEMPIGFEVAFSFLLEIARGINIDVPDPSVLKDIYAKKELKLTIPKIMHK 256
QY 206 MDGYLAIYBEGNLVDMNVKVKYQKMGKSVFNSPSATAAFINQHPGCLNVLNLLDK 265
DB 257 IPTTLLHSLGVRDL-DWEKLLKLQSDGSLFSPSSTAFALMOTKGNCLYLRNNAVKR 315
QY 266 FGNNAVTVPHDLFILSMVDTTIRLGISSHFRVEIKNVLDITYRCWE-----RDEQI 319
DB 316 FNGGVNVPVDFLFEHIVIVDRQLRGISRYFBEELKECLDYVHRITWDNGICWARSCHV 375
QY 320 FMDVVTALAFRLIRLINGEYVSPDLAEITNELAKDEY-----AALETYHAS 367
DB 376 -QDIDDTAMAFRLRQHGQVVSADVFKNPEK-----GEFFCFVQSQNOAVTGMFNLYRAS 430
QY 368 HILYQEDLSGKQILKSADEPLKEIISTDNRLSKLHK-----EVENALKFPINTG 418
DB 431 QLAPPRE-----BILKNAKEFSYNLLEKREBELIDKWIIMKDLPGEIGFALEIPWYAS 485
QY 419 LERINTRNIQLYNVDN-TRILKTTVHSSNITDYLRLAVDFYTCQSIYREBELKGLER 477
DB 486 LPRVETRFYIDQYGGENDVMWIGKTLRYMAYVNNNNLELAKLDYNNCOAHLIENWVIGTWLESGLGFFGL 545
QY 478 WVENKLDQLKFAKQKATCYFSAATLSSPELSDARISWAKNGILTTVVDDFF 531
DB 546 WYENRLSEWGRVSELLECYLAAATIPFSERSHERMVMWAKSVLVKATSSSF 599

RESULT 7

T02959

kaurene synthase A - maize

C:Species: Zea mays (maize)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004

C:Accession: T02959

R:Bensen, R.J.; Johal, G.S.; Crane, V.C.; Tossberg, J.T.; Schnable, P.S.; Meeley, R.B.;

Plant Cell 7, 75-84, 1995

A:Title: Cloning and characterization of the maize Anl gene.

A:Reference number: Z14794; MUID:95210929; PMID:7696880

A:Accession: T02959

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-823 <BEN>

A;Cross-references: UNIPROT:Q41771; EMBL:L37750; NID:g576885; PIDN:AAA73960.1; PID:g5768

C;Genetics:

A;Gene: AN1

C;Function:

A;Description: cyclase involved in gibberellic acid biosynthesis

C;Superfamily: terpene synthase

Query Match 16.0%; Score 657.5; DB 2; Length 823;
Best Local Similarity 30.5%; Pred. No. 2.7e-31;
Matches 192; Conservative 106; Mismatches 266; Indels 65; Gaps 22;

```
QY 12 LTKSRPAALSAIHTASTSHGGTNPNTLIIDTKERIOQKFNKVEISVSYDITAWAMV 71
Db 70 LNRGRKPDLEBHQABEA--ELQP--LIDQVRAML-RSMNDGUTSASADITAWAMV 122
QY 72 P--SPNSPKSPCPPECLNMLNNQNDGSGVLNHTNNHPLK--DSLSTLACIVAL 127
Db 123 PKVGGDGAQPOFPATVRMIVDQLPDGSGW-----DSALFSAYDRMINTLACVVAL 174
QY 128 KRWVVGEDQINKGLGFIESNL-ASATEKSOPSPIGFDIIFPGELLEFAKNDINLLSKQTD 186
Db 175 TKWLEPARCEAGLSFLHNNMWRLAEEASMPIGFEIAPPSLIQTARDLGV-----VD 228
QY 187 F-----SLMLHKLEQKRCCHSNEMDGY---LAVISEGLNGLYDNNMVKYQMKNGSV 236
Db 229 FPGHPALQSIYANREVKLKLPDMWHRVPSIIHLSLEGMPDL-DWPRLLMLQSCDGSF 287
QY 237 FNSPSATAAFAFNHQPGLNVLNLLDKFGNAVPTVYPHDLPFIRLSMVDITIERLGISHH 296
Db 288 LFSPSATAYALMOTGDKCKCFEYIDRVKFGGVENVYVDLFEHIWVVDRLERLIGISRY 347
QY 297 FRVETKVLDEYR-----CWVERDQIFMDVVTICALAFRLIRINGEYSPDPLAEIT 349
Db 348 FQREIQECMDVYNRHWTEDGICWARKSN--VKDVEDDTAMAFKLLHGLNYNPSVFKNFE 405
QY 350 NE-----LALKDEYAALETHY---ASHILYQ-EDLSSGGKILKSADFLKEI-----ISTDS 396
Db 406 KDGEFFCFVGSGTQAVTGMVNLNRASQISFQEDVLHRAVP-SYEFRLQREQGMIRDK 464
QY 397 NRLSKLHKVENALKFPINTGLERINTRNRIQLY-NVDNTRILKTTYHSSNISNTDYL 455
Db 465 WIVAKDLFGEVQYTLDFPWAYSLPRVEARTYLDQYGGKDDVWIGKTLXRMPLVNDNTYLE 524
QY 456 LAVEDFYTCOSIYREELKGLERWVENKLDQKFAKQKTAICYFSVAATLSPPELSDARI 515
Db 525 LAIRDFNHCQALHQECNGLQTYKONCLDAPGVEPDVLRSYFLAAAACIFEPSRAERL 584
QY 516 SWAKNGILTTVVDDFF-DIGGTIDBLTNLIQCVEKWNVDVKDCCSEHVRILFLALKDAI 574
Db 585 AWARTSMIANAISTHLRDI SEDKRLKLECFVHCLYEEN-DVSWLKRNPNDVILERALRLKI 643
QY 575 CWIGDEAFKQW--ARDVTSHVIOQWLELM 601
Db 644 NLLAQEALPIHEGORFIHSLLSLAWTEWM 672
```

RESULT 8

H84633

C;Title: probable limonene cyclase [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C;Accession: H84633

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84633

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-591 <STO>

A;Cross-references: UNIPROT:Q9ZUH4; GB:AE002093; NID:g4115381; PIDN:AA03382.1; GSPDB:GN

C;Genetics:

A;Gene: At2g24210

A;Map position: 2

C;Superfamily: vetispiradiene synthase 1

Query Match 10.9%; Score 446.5; DB 2; Length 591;
Best Local Similarity 27.0%; Pred. No. 5.9e-19;
Matches 138; Conservative 97; Mismatches 242; Indels 35; Gaps 14;

```
QY 281 RLSMVDTIERLIGSHHVRVIEIKNVLDETYRCWVERDEQIEMDVVTICALAFRLIRINGEYV 340
Db 98 QLEFIDDDQKQGVSHFEAEIDNILTSSYK--KDRNTIQESDLHATALEFRFLRQHGFPNV 155
QY 341 SP---DPLAEITNELALKDEYAALETYHSHILYQEDLSSGKQIL---KSADFLKEIIST 394
Db 156 SEDVDFVFMENCGKFPDRDDIYGLISLYEASYL--STKLDKNLQIFIRPATQOOLRDFVDT 213
QY 395 DSNR--LSKLHKVENALKFPINTGLERINTRNRIQLYVNDNTRILKTTYHSSNISNTD 452
Db 214 HSNEDFGSCDMVEIVVQALDMFYQMRRLSTRWIDVYG-----KRONYKNLV 262
QY 453 YLRLLAVEDFYTCOSIYREELKGLERWVENKL-DOLKFAKQKTAICYFSVAATLSPPELS 511
Db 263 VVEFAKIDFNIVQALHQBELKNVSSWMWETGLGKOLYFARDRIVENYFWTIGIQIOBPQYG 322
QY 512 DARISWAKNGILTTVVDDFFDIGGTIDBLTNLIQCVEKWNVDVKDCCSEHVRILFLALK 571
Db 323 YVRQMTKINALTTWIDDIYDYGLEELQLFTVAFENWDIN-RLELPEYNELCFVLVI 381
QY 572 DAICWIGDEAFKQWQARDVTSHVQWLELMNSMLREAIWTRDAYVPTLNEYMENAYVSPA 631
Db 382 NEVNSIACEILRTKNINVIPFLKKSNTDVSXAVLVEAKWYKSHKPNLEENYQNAIRISIS 441
QY 632 LGPIVKPAIYFV-GPKLSEETVESSEYH--NLFKLMSTQGRLLNDIHFKEFEKGLK-N 687
Db 442 -SPITFVHFYCVFSDQLSIQVLETLSSHQNVNRCSSVFRFLANDLVTSPELARGDYCK 500
QY 688 AVALLHLSGSGKVEEVEEVEEMMMKMKRKLKMLKLFEEINGSIVPRACKDAFNMCHVL 747
Db 501 SIQCVMSS--ETGASEDKARSHVRQWINDLWDMYKVAHSSILHHDMEIVINLARMS 558
QY 748 NFFVANDDGFTG---NTILDVTWKDIIYNPLVL 776
Db 559 QCMYQYGDGSGSPEKAKIVDRVMSLLEFNPIEL 590
```

RESULT 9

F71434

C;Title: probable limonene cyclase - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

A;Variety: columbia

C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004

C;Accession: F71434

R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giele

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998

A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,

erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansc

C.; Chaiwalais, N.

A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thali

A;Reference number: A71400; MUID:98121113; PMID:9461215

A;Accession: F71434

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-600 <BEV>

A;Cross-references: UNIPROT:O23516; GB:Z97341; NID:g2244991; PID:g2245028

C;Genetics:

A;Map position: 4COP9-4G3845

C;Superfamily: vetispiradiene synthase 1

Query Match 10.5%; Score 431; DB 2; Length 600;
 Best Local Similarity 24.6%; Pred. No. 5e-16;
 Matches 154; Conservative 116; Mismatches 279; Indels 78; Gaps 17;

QY 180 LLSKOTDPSLML---HKRELEOKRCHSNEMD---GYLAYISEGLNGLYDMNWKYQMK 232
 DB 22 LLSNNTLATLELCHLSLSDAKHCHORNERNORRHGESPSQERLISLIDKNVRLITM 81
 QY 233 NSGVNSPSATAAFINHQPCGLNLSLLDFGNVAVTVYPHDLF--IRLSNVDTIERL 291
 DB 82 SGG-----GVLYKDCGRKEVETAEKAI---LFXELEMLDSLOQL 119
 QY 292 GISHFRVIRKVLDETY--RCWVERDEQIFMDVVTTCALAFRLIRNGYEVSPDPLAET 349
 DB 120 GISYHYKHEIDILAKIHQHGHEIERETO---DLHATSLEFILLRQHGFDVQDAPDVTI 176
 QY 350 NELA-----LKDEYAALFTYHSHILYQEDLSSGKQILKSADFLKEIISTDSNLSK-- 401
 DB 177 SETGEPKTLHSDIKGLLSLYEASYSMDSEFKLRETRIVANKRUSEFVAESSKTICRED 236
 QY 402 --LIHKEVENALKFPINTGLERINTRNRNQLYNDVNTRLKTYTHSSNISNTDYLRLAVE 459
 DB 237 ETVILEMWKRALETYPHWSIRRLERAWYINV-----YEKKHEMNPILLFEAAI 284
 QY 460 DFTQCSIVREELK-----GLERWVENKLDOLKFAKQKTAICYFSVAATLSSPELSDAR 514
 DB 285 DFNMLQANQHEELKLSSTGL-----MKDLDFVRDRITESYFTWIGIFYEPEPKYCR 336
 QY 515 ISWAKNGILTVVDDFDIGTIDELTNLIQCVEKNVND--VDKDCSEHVRILFLALKDA 573
 DB 337 KILTKLFMLIVMDIYDYGTELELELFTNVVEKDVNHER--LPNYMEMCFLYNE 394
 QY 574 ICWIGDEAFKQWQARDVTSVHIQVWLELMSMLREAIWTRDAYVPTLNEYMENAVYFALG 633
 DB 395 INQIGYDVLKDLGNLIPYLKQVWTDLFTFLTESKWKYTKHPSPEEYMQNGVSISSVP 454
 QY 634 PIVKPAIYFVGPKLSEI--IVESSEYHNLFLKMTQGLLNDIHSFKREFKGLNAVALH 692
 DB 455 TILLHLFSLSDHISQTLTDDSKNHSVVRSCATILRLANDLTAISTEEMARGD--SPKSQV 513
 QY 693 LSNESGKVEEVEEVEEMMMKMKELMKLIPEENGSIIVPRACKDAFWNMCHVLNFFYA 752
 DB 514 CYMYETRASEEEARRHQMISDSWDIINSDLKTAHTSSLPGRFLAAAANLRVQCIYR 573
 QY 753 NDGFTG---NTILDVTVDKIYVPLVL 776
 DB 574 HGDGHSPEKTKVTDVYIQSVLFPVPL 600

RESULT 10
 T08174
 sesquiterpene cyclase (BC 2.5.1.1) - pepper
 C;Species: Capsicum annuum (pepper)
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
 R;Back, K.; Shin, D.H.; He, S.
 Plant Cell Physiol. 39, 899-904, 1998
 A;Title: Cloning and bacterial expression of sesquiterpene cyclase, a key branch point e
 A;Reference number: Z16395; MUID:99033462; PMID:9816674
 A;Accession: T08174
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-559 <BAC>
 A;Cross-references: UNIPROT:065323; EMBL:AF061285; NID:g3108342; PIDN:AAC61260.1; PID:g3
 C;Function:
 A;Pathway: the synthesis of phytoalexin capsidiol
 A;Note: unduced by UV
 C;Superfamily: vetispiradiene synthase 1
 C;Keywords: isoprenoid biosynthesis; transferase

Query Match 10.4%; Score 426; DB 2; Length 559;
 Best Local Similarity 25.3%; Pred. No. 9e-18;

Matches 129; Conservative 102; Mismatches 204; Indels 74; Gaps 15;

QY 282 LSNVDTIERLGHSHFRVEIKNVLDIETVRCWVERDEQIF--MDVVTTCALAFRLIRNGYEV 340
 DB 73 LMLIDVIERGLIAYIFEKEIDEILDRINYENSPGSDVYNEEDLCTCRQLRLLRQHGYN 132
 QY 341 SPDPILAETITN-----ELALKDEYAALFTYHSHILYQEDLSSGKQILKSA--DFLKEIIS 393
 DB 133 SLKIFSKFDLNGRLKESLASDVLGLLSLYEASHV-----RSHGEDILDELALAFSTHILE 187
 QY 394 TDSNRLSKLIHKEVENALKFPINTGLERINTRNRNQLYNDVNTRLKTYTHSSNISNTDY 453
 DB 188 SATPHLEYPLKEQVRHALESQSLHKGIPRI-----RIQFF-----ISSYVDKOAIKNVUL 236
 QY 454 LRLAVEDEFTQCSIVREELKGLERWVE--NKLDOLKFAKQKTAICYFSVAATLSSPELSD 512
 DB 237 LRFALKDYNMLQMLHKQELAEVSRWVKDLNFVNTLPYADRVRVCYFWALGVVYEPQYSQ 296
 QY 513 ARISWAKNGILTVVDDFDIGTIDELTNLIQCVEKNVNDKDCSEHVRILFLALKD 572
 DB 297 ARVMLVKTIAMTSIVDDTYDAYGTVDLAIYTDVIQRWDIK--EIDSLPDYMKISYKALLD 355
 QY 573 AICWIGDEAFKQWQARDVTSVHI-----QVWLELMSMLREAIWTRDAYVPTLNEYMENAVY 628
 DB 356 ----LYKDYKEKMSRDSRSHVVYAKERLKLKLVKSYNIEAKWFIEGHMPAPSEYLRNFAV 411
 QY 629 SPALGPIVPAIYFVGPKLSEI-----IVESSEYHNLFLKMTSQGLR 670
 DB 412 TTTYTYIANTS--YLGMKYAKSQQEPWLSKNPKILGCVTICRVIDDIAITYEVEKNRGQL 469
 QY 671 LNDIHSFKREFKGLNAVALHLSNGESGKVEEVEEVEEMMMKMKELMKLIPEENG 730
 DB 470 STGIECYMDYSVSTKEAMAKFOEGESG--WKDINEGML-----RPT 510
 QY 731 IYPRACKDAFWNMCHVLNFFYA--NDGFT 758
 DB 511 PIPMEFLSRILNARLVDVTVYKHNEGYT 539

RESULT 11
 A48863
 limonene cyclase - spearmint
 C;Species: Mentha spicata (spearmint)
 C;Date: 12-May-1995 #sequence_revision 19-May-1995 #text_change 09-Jul-2004
 R;Colby, S.M.; Alonso, W.R.; Katahira, E.J.; McGarvey, D.J.; Croteau, R.
 J. Biol. Chem. 268, 23016-23024, 1993
 A;Title: 4S-limonene synthase from the oil glands of spearmint (Mentha spicata). cDNA in a
 A;Reference number: A48863; MUID:94043077; PMID:8226816
 A;Accession: A48863
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-599 <COL>
 A;Cross-references: UNIPROT:Q40322; GB:L13459; NID:g410229; PIDN:AAC37366.1; PID:g410230
 C;Superfamily: vetispiradiene synthase 1

Query Match 10.0%; Score 413; DB 2; Length 599;
 Best Local Similarity 24.5%; Pred. No. 5.9e-17;
 Matches 135; Conservative 115; Mismatches 209; Indels 92; Gaps 19;

QY 256 LNYNSLLDKFGNAPTVYVPHDLFI-----FLSMVDTIERLGHSHFRVEI 301
 DB 70 VNFQTSLLSDYKEDKHVIRASSELVTLVKMELEKETDQIRQLLELDDQLRGLSDHFOFNEF 129
 QY 302 KNVLDETYRCWVERDEQIF-----MDVVTTCALAFRLIRNGYEVSPDPLAETITNE-- 351
 DB 130 KEILSSIY-----LDHHYKNPFPKEERDLYSTSLAFRLRHEGFQVQAEVDFSKNEEG 184
 QY 352 ---LALKDE--YAALFTYHSHILYQEDLSSGKQILKSA-----DFLKEIIS---TDSNRL 399
 DB 185 EFKESLSDTRGLGLQLYEASFL-----LTEGETTLESAREFATKFLKEKNEGGVDGDL 239
 QY 400 SKLIHKEVENALKFPINTGLERINTRNRNQLYNDVNTRLKTYTHSSNISNTDYLRLAVE 459

```

Db      240  TRIAV-----SLDIPLMRIKR-----FNAPVWIEWYKRPDMNPFWLELAL 282
QY      460  DFTYQTSIVREELKGLERWVVENK-LDOLKEARQKTAYCYPVAATLSPFLSDARISWA 518
Db      283  DLNITVQAQOEELKESFRWRWTGVEKUPFARDRLVECFYNTWTGILIEPRQHASARIMMG 342
QY      519  KNGILTTVVDDFFDGGTIDELTNLIQCVEKNVDVDDKCCSEHVRILFLALKDAICWIG 578
Db      343  KVALITVIDDIYDVVGTLEELQFDTLIRRDIN-SIDQLPDYMQJCFALANNFVDDTS 401
QY      579  DEAFQWQARDTSHVITQWLELMNMLREAIWTRDAYVPTLNMEYNAVVSFALGPIVXP 638
Db      402  YDVMEKGVNVPYLRQSWVDLADKYMVEARFYGCHKPSELYLENSWQSIS-GFCMLT 460
QY      639  AIYF-VGPKLSEIIVES-SEYHNLFKLMSTQGLRLNDIHSFKREFKGL-NAVALHLSN 695
Db      461  HIFRVVTSFKEYETVDSLYKYHDLVRWSFVURLADDIGTSVEEVRGDPVKSLOCYMSD 520
QY      696  GESGKVEEVEEEMMMIKNRKELMKLIFEENGSI-VPRACKDAFW-----NMCH 745
Db      521  YNASEAE-----ARKHKWLLIAEYVKKMAERYSKDSPPFGKDFGICAVDILGR 567
QY      746  VLNPFYANDGG 756
Db      568  MAQLMYHNGDG 578

```

```

RESULT 12
G71434
Probable limonene cyclase - Arabidopsis thaliana
A:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004
C:Accession: G71434
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: G71434
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1024 <BEV>
A:Cross-references: UNIPROT:O23517; GB:Z97341; NID:g2244991; PID:e327011; PID:g2245029
C:Genetics:
A:Map position: 4COP9-4G3845

```

```

Query Match          9.9%; Score 406; DB 2; Length 1024;
Best Local Similarity 25.3%; Pred. No. 3.5e-16;
Matches 141; Conservative 104; Mismatches 226; Indels 86; Gaps 18;

QY      278  LFIRLSMDTIERLIGISHHFRVEIKNVLDIFY-----RCWVER-DEQIFMDVVTTCALAFRL 332
Db      27  LLEQLELIDTLQRLGVSHFEQEIKTTLNVHVKNVRAHKNRINDRNGDLIYATALEFRL 86
QY      333  LRINGEVSDDPLAEINELALKDVEALTYHSHILYQEDLSSGKOILKSADFLKEII 392
Db      87  LR--QHDVFD---GNIGVLDLDDKIKGILSLYEASYLSTRIDTKLESIVYTKRLAKFV 141
QY      393  STDSNRL-SKLHKEVENALKFPINTGLERINFRNTQLYVNDNTRILKTYTHSSNLSNT 451
Db      142  EVKNETKSYTLRKNWVTHALEMPYHRVGRLEARWYIEVYG-----ERHDMNP 189
QY      452  DYLRAVEDFTYTCOSIYREELKGLERWVENKLDQ-LKEARQKTAYCYFVSVAATLSSPEL 510
Db      190  ILLELAKLDFNFVQAIHQDEKLSLSSWSWKTGLTKHLDVDRDITRTEGYFSSVGVMVEPEF 249
QY      511  SDARISWAKNGILTTVVDDFFDGGTIDELTNLIQCVEKNVDVDDKCCSEHVRILFLAL 570

```

```

Db      250  AYHRQMLTKVFMILITIIDDIYIGTLEELQFTTIVEKWDVNRLEE-LPNYMKGLCFCL 308
QY      571  KDAICWIGDEAFKQWQARDVTSHTVITQWLELMNMLREAIWTRDAYVPTLNMEYNAVVSF 630
Db      309  VNEINQIGVFLDRDKGFNVIPYLKESWADMCITFLKEAKWYKSGYKPNFEEYQWQWISS 368
QY      631  ALGPVTKPAIYFVGPKLSEIIVE--SSEYHNLFKLMSTQGLRLNDI-----HSP----K 678
Db      369  SV-PTILHLFCL--LSDQTLDIILGSYNSHVSRSSATILRLANDLATSSVSHGFTYNT 424
QY      679  REFKEG-KLNAVALHLSNGESGKVEEVEEEMMMIKNRKELMKL1-----FEENG 729
Db      425  BELARGDTMKSVQCHWH--ETGASEAE-----SRAYIQIIGLVAMDLDLNNKKKS 471
QY      730  STVPRACKDAFWNMCHVLNFFYANDDG-----FTGNTILDTV 766
Db      472  CRLHQGFLEAAANLGRVAQCVYQGDGHCPCDKAKTVNHRQRIRKWRNAFLGSGIMTTI 531
QY      767  KDIIYNPLVLYNENEBOQ 783
Db      532  VLLHTPTFTFSVFSDEEE 548

```

```

RESULT 13
T03714
5-epi-aristolochene synthase - common tobacco
N:Alternate names: sesquiterpene cyclase
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C:Accession: T03714
R:Facchini, P.J.; Chappell, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 11088-11092, 1992
A:Title: Gene family for an elicitor-induced sesquiterpene cyclase in tobacco.
A:Reference number: Z15024; MUID:93066390; PMID:1438319
A:Accession: T03714
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-550 <FAC>
A:Cross-references: UNIPROT:Q40577; EMBL:L04680; NID:gi70342; PID:AAA19216.1; PID:g50551
A:Experimental source: strain NK326
C:Genetics:
A:Introns: 37/3; 126/1; 251/2; 324/2; 370/3; 452/3
C:Function:
A:Description: mediates the conversion of the isoprenoid intermediate farnesyl diphosphat
C:Superfamily: vetispiradiene synthase 1

```

```

Query Match          9.7%; Score 399; DB 2; Length 550;
Best Local Similarity 23.7%; Pred. No. 3.6e-16;
Matches 123; Conservative 109; Mismatches 216; Indels 70; Gaps 16;

QY      282  LSWVDTIERLIGISHHFRVEIKNVLDIFYRCWVERDEQIFMDVVTTCALAFRLRINGVEVS 341
Db      70  LNLIDTIERLIGISVHFEKIDEILDQIY-----NQNSCNDLCTSAQLRLLRQHGFINIS 124
QY      342  PDPLAEITNE-----LALKDEYAALETTHASHILYQEDLSSGKOILKSADFLKEI-IS 393
Db      125  PEIIFSKPDENGKFKESLA-SDVGLGLNLYEASHVRTHAD-----DILEDALAFSTHLE 178
QY      394  TDSNRLSKLHKEVENALKFPINTGLERINFRNTQLYVNDNTRILKTYTHSSNLSNTDY 453
Db      179  SAAPHLKSPREQVTHALEQCLHKGVPRVETR-----FFISSIYDKQSQKNNVL 227
QY      454  LRJAVEDFTYTCOSIYREELKGLERWVE-NKLDQLKPARQKTAYCYFVSVAATLSSPELSD 512
Db      228  LRPAKLDENLQMLHQLAQLAQSRRWKDLDFVTTLPVARDRVVECFWALGVFFEQYSQ 287
QY      513  ARISWAKNGILTTVVDDFFDGGTIDELTNLIQCVEKNVDVDDKCCSEHVRILFLALKD 572
Db      288  ARVMLVKTISMISIVDDTDFDAYGTVKELEYDAITQWDLIN-EIDRLPDVWKISYKAILD 346
QY      573  AICWIGDEAFKQWQARDVT---SHVITQWLELMNMLR-----EAIWTRDAYVPTLNMEY 624

```

```
Db 347 -----LYKDYEKELSSAGSHIVCHAIERMKVEVRYNINVBSTWTFIEGYMPVSEYLS 398
Qy 625 NAYVSFALGPVVKPAIYFVGPKLSEIIVESSYHNLFKLMSTQGRLLNDIHSFKRFBKGB 684
Db 399 NALATTTVYLLATTVGLGMSKATEQDFEWSKNPKILASVILCRVIDDTATYVEKSRG 458
Qy 685 KLNVALHLSNGESGKVEEVEVMMMLKKNRKLMLKLIPEENGSIIVPRACKDAF--- 740
Db 459 QT-ATGIECCMRDYGISTKEMAKQNMABTAWKDI-----NEGLLRPTPVSTFLTPI 511
Qy 741 WNMCHVLNFFYA-NDGFTG-----NTILDTVK 767
Db 512 LMLARIVEVYTHNLDGTHPEKVLKPHIINLLVDSIK 549

RESULT 14
A:Species: Hyoscyamus muticus
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C:Accession: A56118
R:Back, K.; Chappell, J.
J. Biol. Chem. 270, 7375-7381, 1995
A:Title: Cloning and bacterial expression of a sesquiterpene cyclase from Hyoscyamus muticus
A:Reference number: A56118; MUID:95221394; PMID:7706281
A:Accession: A56118
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-520 <BA>
A:Cross-references: UNIPROT:Q39978; GB:U20188; NID:g763422; PIDN:AAA86337.1; PID:g763423
C:Superfamily: vetispiradiene synthase 1

Query Match 9.7%; Score 398.5; DB 2; Length 520;
Best Local Similarity 23.9%; Pred. No. 3.5e-16;
Matches 127; Conservative 115; Mismatches 224; Indels 65; Gaps 15;

Qy 260 NSLLDFGNAVPIVYPH-----DLFIQLSVMDTIERIGISHHFRVEIKNVLD 308
Db 3 NOVAEKYAEIETLKQGTSTMLSAAGTTITEKINLIDIERIGIAHYHEKQIEDMLDHI 62
Qy 309 YRCWVERDEQIFMDVVTICALAFRLRLRINGVEYDFPLAEITN-----ELAKDEVAAL 362
Db 63 YRADPVEAHEYNDLNTSSVQFLLRHQGVNSENIFSRFQDANGKFKESLRSDIRGLLN 122
Qy 363 TYHASHI-LYQEDLSSGKQLKADFLKEIISDLSNRSLKLIHKEVENALKFPINTGLER 421
Db 123 LYEAHVTRTHKEDI-----LEEALVFSVGHLESAAPHLKSPLSKVTHALEQSLHKSIPR 177
Qy 422 INTRNIQLYNDVTRILKTYTHSSNISNTDYLRLAVEDFYTCQSTYREELKGLERWVVE 481
Db 178 VEIRYFISI-----YEEEFYNDLLRLFAKLDYNIQLMLHKLHLSSEVSRWKD 225
Qy 482 -NKLDQLKFAKQRTAYCYFSAATLSSPELSDAISWAKNGIILTVVDDPFDIGGTIDEL 540
Db 226 LDFVTLLPYARDRAVECYFTWGVYAEQYSQARVMLAKTIAMISIVDDTFDAYGIVKEL 285
Qy 541 TNLIOQVEKNVDVDDCCSEHVRILFLAKDAICWIGDEAFKQWQADVTSHVIQWLEL 600
Db 286 EVYTDATQRWDIS-QIDRLPEYMKISYKALLDLYDDYEKLSKDGSDGVVHYAKRKKEI 344
Qy 601 MNSMLREAIWTRDAYVPTLNEYMENAVVFA-----LGPVVKPAIYFVGPKLSEI 651
Db 345 VGNVFIQKFIQGYMSEVSEIYLSNALTATSYLLITTSYLGMSKATKEHEFWLATNPKI 404
Qy 652 VESSEYHNLFKLMS TOGRLLNDIHSFKRFBKGBL-NAVALHLSN-GESEKVEEVEVVEE 709
Db 405 LEAN-----ATLCRVVDDIATYVEKRGQIATGIECYMRDYGVSTEVAMEKPFQEM 455
Qy 710 M-MMKKNRKLMLKLIPEENGSIIVPRACKDAFNMCHVLNFFYA-NDGFT 758
Db 456 ADIAWKQVNEILRPT-FVSSEILTR-----ILNLARIIDVTYKHNDQGYT 500
```

```
RESULT 15
S68366
A:Species: Gossypium arboreum
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S68366
R:Chen, X.Y.; Chen, Y.; Heinstein, P.; Davisson, V.J.
Arch. Biochem. Biophys. 324, 255-266, 1995
A:Title: Cloning, expression, and characterization of (+)-delta-cadinene synthase: a cat-
A:Reference number: S68365; MUID:96132653; PMID:8554317
A:Accession: S68366
A:Molecule type: mRNA
A:Residues: 1-554 <CH>
A:Cross-references: UNIPROT:Q39760; EMBL:U23205; NID:g1045313; PIDN:AAA93065.1; PID:g104
A:Experimental source: cultivar Nanking
C:Superfamily: vetispiradiene synthase 1
C:Keywords: phytoalexin biosynthesis

Query Match 9.1%; Score 373; DB 2; Length 554;
Best Local Similarity 24.6%; Pred. No. 1.3e-14;
Matches 131; Conservative 98; Mismatches 217; Indels 86; Gaps 20;

Qy 281 RLSDVDTTIERIGISHHFRVEIKNVLDFTYRCWVERDEQIFMDVVTICALAFRLRLRINGYEV 340
Db 73 KLAIFDSVQRLGVSYHFTKEIELENIY-----HNNDAENDLYTTSLEFRLLRHEGFNV 128
Qy 341 SPDLAEITNE-----LALKDEVAALTYHSHIILYQEDLSSGKQLKLSADFLKEIIST 394
Db 129 SCDVFNKFKDEQGNPKSSVTSVDRGLLELYQASLYR-----VHGEDIIDEAISF 177
Qy 395 DSNRLSKLI-----HKEVENALKFPINTGLERINTRRNIQLYNDVTRILKTYTHSSN 447
Db 178 TTNHLSLAVASLDYPLSEVSHALKQSIIRGLPRVEARHYL-----SVYQDIE 225
Qy 448 ISNTDYLRLAVEDFYTCQSIYREELKGLERWVVENKLD--OLKFAKQRTAYCYFSAAT 504
Db 226 SHNKVLLEFAKIDFNVMVQLLHKKELSEISRKKWD--LDFQKLPYARDRVVEGYFWISGV 283
Qy 505 LSSPELSDAISWAKNGIILTVVDDPFDIGGTIDELTNLIQCVERKNVD-VDXDCCSEHV 563
Db 284 YFEPQYSLGRKMLTKVIAMASIVDDYDSYATVEELIPYTKALERWDIKIDE--LPEYM 341
Qy 564 RILFLALKDAICWIGDEAFKQWQADVTSHVIQ-----TWLELMNSMLREAIWTRDAYVPTL 619
Db 342 KPSYKALLD---VYEEMEQLVAKHGRQYREYAKNAMIRLAQSYLVEARWTLQNYKPSF 397
Qy 620 NEYMENA-----YVSFA-----LGPVVKPAIYFVGPKLSEIIVESSEVENLFLKLMSTQ 668
Db 398 EEFKANALPTCGYAMLAITSFVGWGDIVTPTET-FKWAANDPKLIQAS-----TILC 447
Qy 669 RLILNDI-HSPKREFKEGKLNAVALHLSNGESGKVEEVEVVEEVMNMIKKNRKLMLKLIPE 726
Db 448 RFMDDVAEHKFKR-REDDCSALECYME--EYGVTAQAEYDVFNKGVESAWKDVNKEFLK 504
Qy 727 ENGSIVPRACKDAFNMCHVLNFFYANDDQGT--GNWILDTVDKDIYNPLVL 776
Db 505 P--TEMPTVELNRSNLARVMDVLYREGDGYTVVGKAAKGITSLTLEPVAL 554
```

Search completed: October 17, 2004, 08:54:21
Job time : 53.1457 secs

16 NRPAALSAIHTASTSHGQTNPNLIIDT-----TKERQKQFKNVEISVSSYD 64


```
Db 659 VSRDLDLRLVLRKE-SAVPRPKCKLEFKWCKILHLFVSQNDGFSPEKVMYSAVNAVINEP 717
QY 774 LVLVN 778
Db 718 LKVN 722

RESULT 3
US-09-398-395A-56
; Sequence 56, Application US/09398395A
; Patent No. 6468772
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 64687721, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/398,395A
; CURRENT FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-398-395A-56

Query Match 25.8%; Score 1062.5; DB 4; Length 868;
Best Local Similarity 32.3%; Pred. No. 3.5e-85;
Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;

QY 7 IASPLLTNSRPAAL---SAHTASTSGGQNPINLIIDTKERIQKQKNV---EISV 60
Db 79 LSSSTLVKREFPPGFWKDDLDLSLTSKVAASDEKR-IETLISEIKNFRCMGGETNP 137
QY 61 SSYDTAWAMVPSPNSPKSPCPPECLNWLINQLNDGSLVNHHTNHNHPLKDLGSLST 120
Db 138 SAYDTAWARIPAVGSDNHPETVEWILQNLKXGSGW-----EGFYFLAYDRILAT 191
QY 121 LACIVALKRWNVGSDQNKGLSFIESNLASATEKSP-SPIGFDIIPFGLLEYAKNLIN 179
Db 192 LACITTLTWRGTQVQKGIETFRTOAGKMEDEADSHRPSGFETVFPAMLKEAKILGID 251
QY 180 LLSKQTDPSLMLHKELEKQKCHSN---EMDGYLAVISEGLNLYDMNVKVKYQKNGSV 236
Db 252 LPYDLPFLKQIIEKREAKLRIPDVLVLPALPTLLYSLGQEIWDQKIMKLOKSGSF 311
QY 237 FNSPSATAAATINHQNPCLNLYNSLLDFGNAPTVPYPHDIFIRLSMVDITIERIGISH 296
Db 312 LSSPASTAAVEMRTGNKKCLDFNLVKKFGNHVECHYPDLDFELWAVDIVERLIGDRH 371
QY 297 FRVEIKNVLDITYRCWVERD-----EQIFMDVVTCALAFRLRLRINGVESPDP LAETINE 351
Db 372 FKEEIKALDYVYSHWDERGIGWARENPVPDIDDTAMGLRILRLHGVNVSSDLTKTFRDE 431
QY 352 LALKDEYAA-----LETHASHILYQEDLSGKQILKSA-----DFLKE 390
Db 432 ---NGEFCFCFGLQGTORGVTMDLNVNRCSHVSP-----PCETIMEEAKLCTERYLENALEN 483
QY 391 IISTDSNRLSLIKHKEVENALKFPINTGLERINTRRNIOLYNDNTRILKTTIYHSSNIN 450
Db 484 VDAFDKAWFKKNIRGEVEYALKYPHKSKMPLREARSYIENYGPDDVWLGKIVYMPYISN 543
QY 451 TDYLR LAVEDPYTCOSIVREBELKGLERWVVENKLDQLKFAKQTAICYFSAATLSSPEL 510
```

```
Db 544 EKYLELAKLDENKVKQSIHQTELQDLRRWKKSSGFTDLNTRERAVTETIYSPSPAFIEPEP 603
QY 511 SDARISWAKNGILTTVVDDFFDIGGFDIDELTNLIQCVERKNVDVDKDCOCESEHVRILFLAL 570
Db 604 SKREVIYTKTSNFTVILDDLYDAHGSLLDKLFTESVKRWDLSL-VDQMPQOMKICFVG 662
QY 571 KDAICWIGDEAFKQWQARDVTSVIQTLWELMNSMLREALWTRDAYVPTLNEYMENAYVSF 630
Db 663 YNTFNDIAKEGRQRGRDVLGIQNVKVLQRAYTKEAEWSEAKYVPSFNEYIENASVSI 722
QY 631 ALGPIVVKPAIYFVGPKLSEIEVESSEYHNLF-KLMSQTQGRLLNDIHSFKREPKRGKLNAY 689
Db 723 ALGTVVLISALFTGEVLTDEVLSKIDRESRFIQLMGLTGLVNDTXYQARQOGSV-AS 781
QY 690 ALHLSNGESGKV-BEEVVEEMMMKINKKELMKLIFENGSIIVPRACKDAFWNMCHVLN 748
Db 782 AIQCYMKDHPKISEEALQHVYSVMENALEELN---EFVNKIPDIYKRLVFPETARIMQ 838
QY 749 FFYANDDGFT---GNTILDTVKDIIVNPL 774
Db 839 LFYMQGDGLTSLHDMEIKEHVXNCLFPQV 867

RESULT 4
US-09-887-586A-56
; Sequence 56, Application US/09887586A
; Patent No. 6495354
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 64953541, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/887,586A
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-887-586A-56

Query Match 25.8%; Score 1062.5; DB 4; Length 868;
Best Local Similarity 32.3%; Pred. No. 3.5e-85;
Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;

QY 7 IASPLLTNSRPAAL---SAHTASTSGGQNPINLIIDTKERIQKQKNV---EISV 60
Db 79 LSSSTLVKREFPPGFWKDDLDLSLTSKVAASDEKR-IETLISEIKNFRCMGGETNP 137
QY 61 SSYDTAWAMVPSPNSPKSPCPPECLNWLINQLNDGSLVNHHTNHNHPLKDLGSLST 120
Db 138 SAYDTAWARIPAVGSDNHPETVEWILQNLKXGSGW-----EGFYFLAYDRILAT 191
QY 121 LACIVALKRWNVGSDQNKGLSFIESNLASATEKSP-SPIGFDIIPFGLLEYAKNLIN 179
Db 192 LACITTLTWRGTQVQKGIETFRTOAGKMEDEADSHRPSGFETVFPAMLKEAKILGID 251
QY 180 LLSKQTDPSLMLHKELEKQKCHSN---EMDGYLAVISEGLNLYDMNVKVKYQKNGSV 236
Db 252 LPYDLPFLKQIIEKREAKLRIPDVLVLPALPTLLYSLGQEIWDQKIMKLOKSGSF 311
QY 237 FNSPSATAAATINHQNPCLNLYNSLLDFGNAPTVPYPHDIFIRLSMVDITIERIGISH 296
Db 312 LSSPASTAAVEMRTGNKKCLDFNLVKKFGNHVECHYPDLDFELWAVDIVERLIGDRH 371
```


| | | | |
|--|-----|--|-----|
| Db | 723 | ALGTVLISALTFGEVLTDEVLSKIDRESRFQLMGLTGRVNDYTKYQAEQGQEV-AS | 781 |
| Qy | 690 | ALHLSNGESGKV-EEVVEEMMMIKKVKMLKLIFFENGSGIVPRACKDAFNMCHVLN | 748 |
| Db | 782 | AIQYMKDHPKISEEBALQHVSVMENALEELNR---EFVNNKIPDIYKRLVFETARIMQ | 838 |
| Qy | 749 | FFYANDDGT---GNVILTVDKIYNPL | 774 |
| Db | 839 | LFYMQDGLTLGSHDMEIKHVKRNCLPOFV | 867 |
| RESULT 8 | | | |
| US-09-315-861-2 | | | |
| ; Sequence 2, Application US/09315861 | | | |
| ; Patent No. 6114160 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Rodney B. Croteau, Mark R. Wildung | | | |
| ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL | | | |
| ; TITLE OF INVENTION: BIOSYNTHESIS | | | |
| ; NUMBER OF SEQUENCES: 6 | | | |
| ; CORRESPONDENCE ADDRESS: | | | |
| ; ADDRESSEE: Donald L. Stephens Jr. | | | |
| ; STREET: One World Trade Center | | | |
| ; STREET: 121 S.W. Salmon Street | | | |
| ; STREET: Suite 1600 | | | |
| ; CITY: Portland | | | |
| ; STATE: Oregon | | | |
| ; COUNTRY: United States of America | | | |
| ; ZIP: 97204 | | | |
| ; COMPUTER READABLE FORM: | | | |
| ; MEDIUM TYPE: Disk, 3-1/2 inch | | | |
| ; COMPUTER: IBM PC compatible | | | |
| ; OPERATING SYSTEM: MS DOS | | | |
| ; SOFTWARE: Wordperfect 5.1 | | | |
| ; CURRENT APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US/09/315,861 | | | |
| ; FILING DATE: | | | |
| ; CLASSIFICATION: | | | |
| ; PRIOR APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: 08/843,363 | | | |
| ; FILING DATE: | | | |
| ; ATTORNEY/AGENT INFORMATION: | | | |
| ; NAME: Donald L. Stephens Jr. | | | |
| ; REGISTRATION NUMBER: 34,022 | | | |
| ; REFERENCE/DOCKET NUMBER: 4630-46842/DLS | | | |
| ; TELECOMMUNICATION INFORMATION: | | | |
| ; TELEPHONE: (503) 226-7391 | | | |
| ; TELEFAX: (503) 228-9446 | | | |
| ; INFORMATION FOR SEQ ID NO: 2: | | | |
| ; LENGTH: 862 amino acid residues | | | |
| ; TYPE: amino acid | | | |
| ; STRANDEDNESS: single stranded | | | |
| ; TOPOLOGY: linear | | | |
| US-09-315-861-2 | | | |
| Query Match 24.6%; Score 1013.5; DB 3; Length 862; | | | |
| Best Local Similarity 32.2%; Pred. No. 7.9e-81; | | | |
| Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16; | | | |
| Qy | 57 | EISVSYDTAWAVP--SPNSPKSPCFPECLNWLINNQLNDGSGVLNHNHPLLK | 114 |
| Db | 134 | DISPASYDTAWARLATISDGEKFPFQALNWFNNQLDGSWGIESHFS-----LC | 187 |
| Qy | 115 | DSLSTLACIVALKRWNVGSDQINKLSGFIESNLASATEKSQSPISGDIIFPGLLYAK | 174 |
| Db | 188 | DRLLNTTNSVIALSWKVTGHSQVQGAETFAENLRLLNEDELSP-DFQIIFPALLOKAK | 246 |
| Qy | 175 | NLDNLILSKQDFSLMLHKB--LEQKRCHSNEMDGLYLAISGLGNLYDNWNVKQYQM | 232 |
| Db | 247 | ALGINLPDLPFIKYLSTTREARLTVSAAADNPANMLNLEGLEEVIDWNKINRFQSK | 306 |
| Qy | 233 | NGSVFNPSATAAFINHQNPGLNLYNSLLDKFGNAVETVYPHDLFIRLSMVDFTIERLG | 292 |
| Db | 307 | DGSLFSSPASTACVLMTGDKCFTFLNNLLDKFGGCVPCWYSIDLLERLSLVNDNIHLG | 366 |
| Qy | 293 | ISHFRVEIKNVLDYTRCWVERD-----EQFMDVVVTCALAFRLLRINGEVSDDPLAE | 347 |
| Db | 367 | IGRHFKQEIKGALDYVYRHWSEGIWGGRDSLVDPDNTTALGLRTRHMGYNVSSVLNN | 426 |
| Qy | 348 | ITNELALKDE-----YAALETWHASHILYQEDISGKQLILKSADFLEKI | 392 |
| Db | 427 | -----FKDENGFRFSSAQGTHVELRSVNVLPASDLAFDPERAMDDARKFAEPYLREAL | 480 |
| Qy | 393 | STDNSRLSLIHKVEENALKFPINTGLERINTRNQLYNVDNTRILKTYTHSSNISNTD | 452 |
| Db | 481 | ATKISNTKLF-KEIEYVVEYDWMGSIPLREARSYIDSYDDNVVMQKTLYRMPSLNSK | 539 |
| Qy | 453 | YLRAVEDPYTCOSIYREBELKGLERVVENKLDLQKFAQKTAYCYFSVAATLSSPELSD | 512 |
| Db | 540 | CLEAKLDFNIVQSLHQEELKLLTRWKSADMINTFRRVAEVFSSAT--FEPEYSA | 597 |
| Qy | 513 | ARISWAKNGILTTVDDDFDIDGTTIDELTNLIQCCEKKNVDV---DKDCCSEHVRILFLA | 569 |
| Db | 598 | TRIAFTKIGLQVLFDDMADIFATLDELKSFTEGVKSWDTSLHHEIPECMQTCFKVWFKL | 657 |
| Qy | 570 | LKDAICWIGDEAPKQWQARDVTSHVIOTWLELNMMLREAIWTRDAVPTILNEYMENAYVS | 629 |
| Db | 658 | MBE---VNDVVVKQGRDLAHIKRPWELYNFYVQEREWLEAGYIPTFEYLKTYAIS | 713 |
| Qy | 630 | FALGPVTKALYFVGPKLSEEVESSEY-HNLFKLMSTQGRLLNDIHSKREPKEG-KLN | 687 |
| Db | 714 | VGLGPTCTLOPILLMGBLVKDDVVEKVHPSNMPELVSLSWRLTNDTKTYQAEKARGQAS | 773 |
| Qy | 688 | AVALHLSNGESGKVEEVEVVEEMMMIKNRKELMKLIFFENGSGIVPRACKDAFNMCHVL | 747 |
| Db | 774 | GIACYMKD-NPGATEEDAKHICRVVDRAKESFEVFKPSNDI-PWGCKSFIFNLRLCV | 831 |
| Qy | 748 | NFFYANDDGF--TGNLTIDTVKDIYNPL | 774 |
| Db | 832 | QIFYKFDGVIAGIANEBIKDYIRKVIIDPI | 860 |
| RESULT 9 | | | |
| US-09-398-395A-44 | | | |
| ; Sequence 44, Application US/09398395A | | | |
| ; Patent No. 6468772 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Chappell, Joseph | | | |
| ; APPLICANT: No. 64687721, Joseph P. | | | |
| ; APPLICANT: Starks, Courtney M. | | | |
| ; APPLICANT: Manna, Kathleen R. | | | |
| ; TITLE OF INVENTION: SYNTHASES | | | |
| ; FILE REFERENCE: 07678-025001 | | | |
| ; CURRENT APPLICATION NUMBER: US/09/398,395A | | | |
| ; CURRENT FILING DATE: 1999-09-17 | | | |
| ; PRIOR APPLICATION NUMBER: 60/100,993 | | | |
| ; PRIOR FILING DATE: 1998-09-18 | | | |
| ; PRIOR APPLICATION NUMBER: 60/130,628 | | | |
| ; PRIOR FILING DATE: 1999-04-22 | | | |
| ; PRIOR APPLICATION NUMBER: 60/150,262 | | | |
| ; PRIOR FILING DATE: 1999-08-23 | | | |
| ; NUMBER OF SEQ ID NOS: 58 | | | |
| ; SOFTWARE: FastSeq for Windows Version 3.0 | | | |
| ; SEQ ID NO 44 | | | |
| ; LENGTH: 862 | | | |
| ; TYPE: PRT | | | |
| ; ORGANISM: Taxis brevifolia | | | |
| US-09-398-395A-44 | | | |
| Query Match 24.6%; Score 1013.5; DB 4; Length 862; | | | |
| Best Local Similarity 32.2%; Pred. No. 7.9e-81; | | | |
| Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16; | | | |
| Qy | 57 | EISVSYDTAWAVP--SPNSPKSPCFPECLNWLINNQLNDGSGVLNHNHPLLK | 114 |


```

; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/895,752
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxus brevifolia
US-09-895-752-44

Query Match      24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;
Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;

QY 57 EISVSSDYTAWAWVP--SPNSPKSPCFPECLNWLNNQNDGSLGVLNHNHPLK 114
Db 134 DISPSAYDTAWARLATISSDGSEKPRFQALNWFNNQDQSGWGESHSF-----LC 187
QY 115 DLSSTLACIVALKRWNVGDDQNGKLSFIESNLASATEKSQSPGDFIIFGLLEYAK 174
Db 188 DRLNTNSVIALSVKTKHSQVQQAFAENLRLLNEDELSF-DFOIIFPALLQKAK 246
QY 175 NLDINLLSKQTDPSLMLHKE--LEQKRCHSNEMDGYLAIYISGLNGLYDNNMVKYQMK 232
Db 247 ALGINLPYDLPFTKYSTTREAARTDVSAAADNIPANMLNALBGLBEVIDNWKIMRFSQK 306
QY 233 NGSVFNSPSATAAFAINHQNPGCLNVLNLLDKFGNAVPTVYPHDLFIRLSMVDTTIERLG 292
Db 307 DGSFLLSPASTACVLMNTGDEKFTFLNLLDKFGGCVPCWYSIDLLERLSLVNDIEHLG 366
QY 293 ISHFVRVEIKVLDDETVRCWVERD-----EQIFMDVVTCALAFRLARINGEYSPDPLAE 347
Db 367 IGRHFQKEIKGALDYVYRHSERGIGWGRDLSVPDLNTTALGLRTLRLMHGYNVSSDVLNN 426
QY 348 ITNELAKDB-----YAALETYHSHLYQEDLSSGKQILKSADFLKEII 392
Db 427 -----FKDENGGRFFSAGQTHVELRSVNLFRASDLAFPPDERAMDDARKFAEPYLRAL 480
QY 393 STDSNRLSLKHKEVENALKFPINTGLERINTRNITQLYNVNDTRILKTYTHSSNINTD 452
Db 481 ATKISTNTKLF-KIEIYVVEYPWHSIPRLARSYIDSYDNNYVWQKTYLRMPSLNSK 539
QY 453 YRLAIVEDFTYCQSIYREELKGLERWVVENKLDQKFAKQKATACYPFSAATLSSPELSD 512
Db 540 CLELAKLDFNIVQSLHQEELKLLTRWKESGMADINFTHRVAVFPPSSAT--FEPEYSA 597
QY 513 ARISWAKNGILTTVDDDFDGGTIDELTNLIOCEKKNVDV---DKDCCSEHVRILFLA 569
Db 598 TRIAFTKIGCLOVLFDDMAIFATLDKSFTEGVKRWDTSLHHEIPECMTQCFKWWFKL 657
QY 570 LKDAICWIGDEAFKQKQARDVTSVHOTWLELMSMLREAITWRDAYVPTTLNENYENAYS 629
Db 658 MEE-----VNNDVVKVQGRDMLAHIRKPWELYNFCYQEREWLEAGYIPTPEEYLKTYAIS 713
QY 630 FALGPVTKPAIYFGPKLSIEVESSEY-HNLFKLMSTQGRLLNDIHSFKBEFKG-KLN 687
Db 714 VGLGPCVTLQPIILMGLVKKDDVVEKHYPSNMFELSVLSRLNTDITQYQAEKARGOQAS 773
QY 688 AVALHLSNGSGKVEEVEVEMMMIKKREKELMKLIFBENGSIIVPRACKDAFNMCHVL 747
Db 774 GIACYMKD-NPGATEEDAIRKICRVVDRLKESAFEFYFKPSNDI-PWGCKSFIFNLCLV 831
QY 748 NFFYWDGPF--TGNITLTDVKLIYNPL 774

```

```

Db 832 QIFYKFDGCGYIANBEIKDYIRKVIDPI 860

RESULT 12
US-09-903-012B-44
; Sequence 44, Application US/09903012B
; Patent No. 6569656
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65696561, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/903,012B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxus brevifolia
US-09-903-012B-44

Query Match      24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;
Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;

QY 57 EISVSSDYTAWAWVP--SPNSPKSPCFPECLNWLNNQNDGSLGVLNHNHPLK 114
Db 134 DISPSAYDTAWARLATISSDGSEKPRFQALNWFNNQDQSGWGESHSF-----LC 187
QY 115 DLSSTLACIVALKRWNVGDDQNGKLSFIESNLASATEKSQSPGDFIIFGLLEYAK 174
Db 188 DRLNTNSVIALSVKTKHSQVQQAFAENLRLLNEDELSF-DFOIIFPALLQKAK 246
QY 175 NLDINLLSKQTDPSLMLHKE--LEQKRCHSNEMDGYLAIYISGLNGLYDNNMVKYQMK 232
Db 247 ALGINLPYDLPFTKYSTTREAARTDVSAAADNIPANMLNALBGLBEVIDNWKIMRFSQK 306
QY 233 NGSVFNSPSATAAFAINHQNPGCLNVLNLLDKFGNAVPTVYPHDLFIRLSMVDTTIERLG 292
Db 307 DGSFLLSPASTACVLMNTGDEKFTFLNLLDKFGGCVPCWYSIDLLERLSLVNDIEHLG 366
QY 293 ISHFVRVEIKVLDDETVRCWVERD-----EQIFMDVVTCALAFRLARINGEYSPDPLAE 347
Db 367 IGRHFQKEIKGALDYVYRHSERGIGWGRDLSVPDLNTTALGLRTLRLMHGYNVSSDVLNN 426
QY 348 ITNELAKDB-----YAALETYHSHLYQEDLSSGKQILKSADFLKEII 392
Db 427 -----FKDENGGRFFSAGQTHVELRSVNLFRASDLAFPPDERAMDDARKFAEPYLRAL 480
QY 393 STDSNRLSLKHKEVENALKFPINTGLERINTRNITQLYNVNDTRILKTYTHSSNINTD 452
Db 481 ATKISTNTKLF-KIEIYVVEYPWHSIPRLARSYIDSYDNNYVWQKTYLRMPSLNSK 539
QY 453 YRLAIVEDFTYCQSIYREELKGLERWVVENKLDQKFAKQKATACYPFSAATLSSPELSD 512
Db 540 CLELAKLDFNIVQSLHQEELKLLTRWKESGMADINFTHRVAVFPPSSAT--FEPEYSA 597
QY 513 ARISWAKNGILTTVDDDFDGGTIDELTNLIOCEKKNVDV---DKDCCSEHVRILFLA 569
Db 598 TRIAFTKIGCLOVLFDDMAIFATLDKSFTEGVKRWDTSLHHEIPECMTQCFKWWFKL 657
QY 570 LKDAICWIGDEAFKQKQARDVTSVHOTWLELMSMLREAITWRDAYVPTTLNENYENAYS 629
Db 658 MEE-----VNNDVVKVQGRDMLAHIRKPWELYNFCYQEREWLEAGYIPTPEEYLKTYAIS 713

```

QY 630 FALGPVPAIYFVGPKLSEHVESSEY-HNLFLKLMSTQRLINDIHSFKREKFG-KLN 687
D 630 FALGPVPAIYFVGPKLSEHVESSEY-HNLFLKLMSTQRLINDIHSFKREKFG-KLN 687
Db 714 VGLGCTLOPILLMGLVKDDVEKVPNSNFELVSLRWLTNDTKTYQAEKARQQAS 773
QY 688 AVALHLSNGESKVEEVVEENMMKIKRKBELMKLIIFEENGSIYPRACKDAFNMCHYL 747
D 774 GIACYMKD-NPGATBEDAIKHICRVVDRAKLEASFEYFKPSNDI-PMGCKSFIFNLRLCV 831
QY 748 NFFYANDDGF--TGNLTLDVVDIYNPL 774
D 832 QIFYKFDGYGIANEIKDYIRKVIIDPI 860
RESULT 13
US-09-593-253-2
; Sequence 2, Application US/09593253
; Patent No. 6610527
; GENERAL INFORMATION:
; APPLICANT: Rodney B. Croteau, Mark R. Wildung
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL
; BIOSYNTHESIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Donald L. Stephens Jr.
; STREET: One World Trade Center
; 121 S.W. Salmon Street
; Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/593,253
; FILING DATE: 13-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/843,363
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Donald L. Stephens Jr.
; REGISTRATION NUMBER: 34,022
; REFERENCE/DOCKET NUMBER: 4630-46842/DLS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 862 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-593-253-2
Query Match 24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;
Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;
QY 57 EISVSSYDANVAMP--SPNSPKSPCEPCINLNNQNDGSLVNNHNNHPLK 114
D 134 DISPAYTANVARLATISSDGEKPRFPQALNWFNNQNDGSLVNNHNNHPLK 187
QY 115 DLSSTLACIVALKRWNVGDIQNKGLSFIBSNLASATEKSQSPSPIGFDIIFPGILEYAK 174
D 188 DRLNTTNSVIALSVWKGHGVQVQGAFFIENLRLNEDELSF-DFOIIFPALLQAK 246
QY 175 NLDINLASKOTDFSIMLHKRE--LEQKCHSNEMDGYLAIYISGLGNLYDNMVKYQMK 232

Db 247 ALGINLPYDLPIKXLTREARLTDVSAADNIIPANMLNALEGLEEVDMKWKRFQSK 306
QY 233 NGSVNPSATAAAFINQNPCLNYANSLDKFGNAVTVYPHDLFIRLSMVDTIERLG 292
D 307 DGSFLSSPASTACVLMNTGDEKCFITLNLDDKFGCVPCMYSIDLLERLSVDNIEHLG 366
QY 293 ISHFERVEIKNVLDITYRCWVERD-----EQIFMDVVTICALAFRLIRINGIYVSDPLAE 347
D 367 IGRHFQKEIKGALDYVYRHWSEGIWGGRDLSVLPDLNTTALGLRTHRMHGYNNVSDVLNN 426
QY 348 ITNELALKDE-----YAALETYHSHILYQEDLSSGKQLTKSADFLKEII 392
D 427 -----FKDNGRFFSSAGQTHVELRSVNLPRASDLAPPDERAMDARKFAEPVIREAL 480
QY 393 STDSNRLSKLHKVEENALKPPINTGLERINTRNIQLYNVNDNTRILTKTYHSSNISNTD 452
D 481 ATKISTNTKLF-KELEYVVEYEPHWSIPLREARSYIDSYDDNVMQRTLYRMPISLSNSK 539
QY 453 YLRLAVEDFYTCOSIYREELKGLERWVVENKLDQLKEAROKTAYCVFSVAATLSSPELSD 512
D 540 CLELAKLDFNIVQSLHQEELKLLTRWKESGADINFTRHVAEYVFSAT--FPEYSA 597
QY 513 ARISWAKGILTVVDDFFDIGTIDELTNLIQCVEKNVDY--DXDCCSEHVRILFLA 569
D 598 TRIAFTKIGCLQVLPDDMADIFATLDELKSFTEGVKRWDTSLHLEIPECWQCFKVFEL 657
QY 570 LKDALCWTGDEAFKQWQARDVTSHTVQTWLELMNSMLRAIWRDAYVPTLNEYMNAYVS 629
D 658 MEE-----VNDVVKVQGRDMLAHIRKPMELVFCVQVQREMLEAGYIPTFEYLKTYAIS 713
QY 630 FALGPVPAIYFVGPKLSEHVESSEY-HNLFLKLMSTQRLINDIHSFKREKFG-KLN 687
D 714 VGLGCTLOPILLMGLVKDDVEKVPNSNFELVSLRWLTNDTKTYQAEKARQQAS 773
QY 688 AVALHLSNGESKVEEVVEENMMKIKRKBELMKLIIFEENGSIYPRACKDAFNMCHYL 747
D 774 GIACYMKD-NPGATEEDAIIKHICRVVDRAKLEASFEYFKPSNDI-PMGCKSFIFNLRLCV 831
QY 748 NFFYANDDGF--TGNLTLDVVDIYNPL 774
D 832 QIFYKFDGYGIANEIKDYIRKVIIDPI 860
RESULT 14
US-09-900-797-44
; Sequence 44, Application US/09900797
; Patent No. 6645762
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No 66457621, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/900,797
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxus brevifolia
US-09-900-797-44
Query Match 24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;

| Query Match | 23.58; | Score 965; | DB 4; | Length 782; |
|-----------------------|--------|---|-----------------|---------------------|
| Best local Similarity | 31.86; | Pred. No. 1.46-76; | | |
| Matches | 240; | Conservative 145; | Mismatches 318; | Indels 52; Gaps 15; |
| Qy | 58 | ISVSSYDTAMVAVSPNSPKSPCFPECLNWLINQLNDGSGWGLWHTHHNHPLLKDSL | 117 | |
| Db | 44 | ITPSAYDTAMVAVPAIDGSGARPPQPTVDWLKNQLKDSWGI-----QSHFLSDRL | 97 | |
| Qy | 118 | SSTLACTIVALKRMNVGBDQINKGSLFTIESNLASATEKSQSPPI--GFDIIPPGLLLEVANN | 175 | |
| Db | 98 | LATLSCVLVLLKMNVBGLQVEQIEFTIKSNLELVKDETQDSLVTDFEIIIPPSLLREAGS | 157 | |
| Qy | 176 | LDINLLSKQDTFSLMLHKEBLEOKRCHSNE---MDGYLAYISEGLNLYDMNMVKKYQMK | 232 | |
| Db | 158 | LRGLGVPDLPIYHLQTKRERLAKLGRREIYAVPSLLYSLEGIQDIVERMERIVEQSQ | 211 | |
| Qy | 233 | NGSVNFSFATAAAFINHQHPGCLINYSLLDKFNGNAVPTVPYHDLFIRLSMVDTIERLG | 292 | |
| Db | 218 | DGSLSPASTACVFMETGDAKLEFLNSVMIKFGNVPCCLYPVDLLERLLIIVDNVIRLG | 277 | |
| Qy | 293 | ISHFRVVEIKNVLDDETRCWWERD-----EQIFMDVVTALAFRLLRINGVEVSP--DP | 344 | |
| Db | 278 | IYRHFEXEIKREALDYVYRHHMERIGWGRLNPIADLETLAGFLRLHRYNVSPFAIPDN | 337 | |
| Qy | 345 | LAETITNELAL-----KDEVALETYHASHILYQEDLSSGKOILLKSA-----DPLKEII- | 392 | |
| Db | 338 | PKDANGKPICTSGQFNKDVASMLNLYRASQAF-----PGENILDEAKSPATKYLREALE | 392 | |
| Qy | 393 | -----STDNRLSKLIHKEVENALKFPIINTGLERINTERNIQLYNDVNTRIKTTYHSSN | 447 | |
| Db | 393 | KSETSSAWNKKQN--LSQEIYKALKTSHWASVPRVEAKRYQCVYRPDVARIAKCYKLPY | 450 | |
| Qy | 448 | ISNTDYLRILAVEDPYTQSIYREELKGLRRVVENKLDLKFAKQKTAICYFSVAATLSS | 507 | |
| Db | 451 | VNNEKFTLGLKLDENIIQSIHOEEKNKMTSFRDSGLPLFTFARERPLIEFYFLVAAGTYE | 510 | |
| Qy | 508 | PELSDARISWAKNGILLTVVDDFDIGGTIDELTNLIIQCVBEKNWVDKCCSEHVRILF | 567 | |
| Db | 511 | PQYAKRFLFKVACLQTVLDDNDYTYGLDELKLPTEAVRRNDLSFTEN-LPDYMKLCY | 569 | |
| Qy | 568 | LALKDAICWIGDEAKFQKQARDVTSHTVQTLWELMNSMLREAIWTRDAVYPTLMEVNAV | 627 | |
| Db | 570 | QIYYDIIVHEVAWEAKEQGRELVSFRRKGWEDYLLIGYVEEAELAAEYVPTLDEYIKNGI | 629 | |
| Qy | 628 | VSPALGPVTKPAIYFV-GPKLSEELVESSEVHN--LFLKLMSTOGRLANDIHSFKREFKE | 683 | |
| Db | 630 | TSIGORILLGSLVIMDQQLLSQALEKVDYVPGRRVLTSLNLSITSLRADDTKYKAEAKAR | 689 | |
| Qy | 684 | GKLNVALHLNGBSGKVEEVEBEMMMKIKKELMKLIIFEENGSIYVPPACKDAFWMN | 743 | |
| Db | 690 | GELASSIECYMKDHPCEETEEALDHIYSLSPAVKELTRFLKPD--VPPACKQMLFEE | 747 | |
| Qy | 744 | CHVLNFFVANDGGF--TGNTILDTVVKOIIYNPLVL | 776 | |
| Db | 748 | TRVTWVIFKDGDFGVSKLGVKDHICEILBPL | 782 | |

RESULT 15
US-09-360-545-16
; Sequence 16, Application US/09360545
; Patent No. 6429014
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Bohlmann, Jorg
; APPLICANT: Steele, Christopher L
; APPLICANT: Phillips, Michael A
; TITLE OF INVENTION: MONOTERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; FILE REFERENCE: wslu13885
; CURRENT APPLICATION NUMBER: US/09/360,545
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/052,249
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: PCT/US98/14528
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 107

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:53:05 ; Search time 73.5657 Seconds
(without alignments)
3445.655 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSLCIASPLLTGKSRPAA.....TVKDIYNPLVLNENEQR 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 32331874 residues

Total number of hits satisfying chosen parameters: 1360919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|----------------------|
| 1 | 4113 | 100.0 | 784 | 15 | US-10-041-018-383 |
| 2 | 4026 | 97.9 | 784 | 15 | US-10-041-018-384 |
| 3 | 2855.5 | 69.4 | 788 | 15 | US-10-041-018-390 |
| 4 | 2101.5 | 51.1 | 785 | 15 | US-10-041-018-389 |
| 5 | 2023.5 | 49.2 | 789 | 15 | US-10-041-018-395 |
| 6 | 1675 | 40.7 | 780 | 15 | US-10-425-114-65662 |
| 7 | 1537.5 | 37.4 | 740 | 15 | US-10-259-194A-108 |
| 8 | 1458 | 35.4 | 689 | 16 | US-10-437-963-121564 |
| 9 | 1319 | 32.1 | 770 | 16 | US-10-437-963-138526 |
| 10 | 1273 | 31.0 | 730 | 16 | US-10-437-963-174639 |
| 11 | 1254 | 30.5 | 628 | 15 | US-10-425-114-49083 |
| 12 | 1209 | 29.4 | 1157 | 16 | US-10-437-963-174635 |
| 13 | 1075 | 26.1 | 795 | 13 | US-10-041-007-37 |
| 14 | 1075 | 26.1 | 814 | 13 | US-10-041-007-35 |

| | | | | | | |
|----|--------|------|-----|----|----------------------|-------------------|
| 15 | 1075 | 26.1 | 873 | 13 | US-10-041-007-2 | Sequence 2, Appli |
| 16 | 1075 | 26.1 | 873 | 13 | US-10-041-007-33 | Sequence 33, Appl |
| 17 | 1075 | 26.1 | 873 | 15 | US-10-041-018-398 | Sequence 398, App |
| 18 | 1062.5 | 25.8 | 868 | 9 | US-09-895-752-56 | Sequence 56, Appl |
| 19 | 1062.5 | 25.8 | 868 | 9 | US-09-887-586A-56 | Sequence 56, Appl |
| 20 | 1062.5 | 25.8 | 868 | 9 | US-09-903-012-56 | Sequence 56, Appl |
| 21 | 1062.5 | 25.8 | 868 | 10 | US-09-900-797-56 | Sequence 56, Appl |
| 22 | 1062.5 | 25.8 | 868 | 11 | US-09-893-820-56 | Sequence 56, Appl |
| 23 | 1062.5 | 25.8 | 868 | 11 | US-10-041-007-4 | Sequence 4, Appli |
| 24 | 1062.5 | 25.8 | 868 | 13 | US-10-041-018-388 | Sequence 388, App |
| 25 | 1054 | 25.6 | 746 | 13 | US-10-041-018-39 | Sequence 39, Appl |
| 26 | 1013.5 | 24.6 | 862 | 9 | US-09-895-752-44 | Sequence 44, Appl |
| 27 | 1013.5 | 24.6 | 862 | 9 | US-09-887-586A-44 | Sequence 44, Appl |
| 28 | 1013.5 | 24.6 | 862 | 9 | US-09-903-012-44 | Sequence 44, Appl |
| 29 | 1013.5 | 24.6 | 862 | 10 | US-09-900-797-44 | Sequence 44, Appl |
| 30 | 1013.5 | 24.6 | 862 | 11 | US-09-893-820-44 | Sequence 44, Appl |
| 31 | 1013.5 | 24.6 | 862 | 13 | US-10-041-007-41 | Sequence 41, Appl |
| 32 | 1013.5 | 24.6 | 862 | 15 | US-10-041-018-386 | Sequence 386, App |
| 33 | 965 | 23.5 | 782 | 9 | US-09-895-752-46 | Sequence 46, Appl |
| 34 | 965 | 23.5 | 782 | 9 | US-09-887-586A-46 | Sequence 46, Appl |
| 35 | 965 | 23.5 | 782 | 9 | US-09-903-012-46 | Sequence 46, Appl |
| 36 | 965 | 23.5 | 782 | 10 | US-09-900-797-46 | Sequence 46, Appl |
| 37 | 965 | 23.5 | 782 | 11 | US-09-893-820-46 | Sequence 46, Appl |
| 38 | 965 | 23.5 | 782 | 13 | US-10-041-007-16 | Sequence 16, Appl |
| 39 | 965 | 23.5 | 782 | 14 | US-10-025-145A-16 | Sequence 16, Appl |
| 40 | 965 | 23.5 | 817 | 9 | US-09-865-171-13 | Sequence 13, Appl |
| 41 | 965 | 23.5 | 817 | 9 | US-09-865-171-38 | Sequence 38, Appl |
| 42 | 965 | 23.5 | 817 | 9 | US-09-865-171-40 | Sequence 40, Appl |
| 43 | 965 | 23.5 | 817 | 13 | US-10-041-007-15 | Sequence 15, Appl |
| 44 | 962 | 23.4 | 817 | 9 | US-09-865-171-42 | Sequence 42, Appl |
| 45 | 930 | 22.6 | 470 | 16 | US-10-437-963-196878 | Sequence 196878, |

ALIGNMENTS

RESULT 1
US-10-041-018-383
; Sequence 383, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P. T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041, 018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 383
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Stevia rebaudiana
US-10-041-018-383

Query Match 100.0%; Score 4113; DB 15; Length 784;
Best Local Similarity 100.0%; Pred. No. 8.9e-298;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | | | |
|----|-----|----------------------|--------------------|---------------------|-----|
| QY | 1 | MNLSLCIASPLLTGKSRPAA | SAIHTASTSGGQNPNTLI | DTTKRIQKQKNVEISV | 60 |
| DB | 1 | MNLSLCIASPLLTGKSRPAA | SAIHTASTSGGQNPNTLI | DTTKRIQKQKNVEISV | 60 |
| QY | 61 | SSYDTAWAMVSPNSPKSP | CFPECLNWLNNQNDGSG | VLNTHNNHPLKDSLSST | 120 |
| DB | 61 | SSYDTAWAMVSPNSPKSP | CFPECLNWLNNQNDGSG | VLNTHNNHPLKDSLSST | 120 |
| QY | 121 | LACIVALKRWNVGPDQINK | GLSFIESNLASATEKSQ | SPISGDIIFPGLLYAKNLD | 180 |
| DB | 121 | LACIVALKRWNVGPDQINK | GLSFIESNLASATEKSQ | SPISGDIIFPGLLYAKNLD | 180 |

181 LSKOTDPSMLHKELEOKRCHSNEMDGYLAYSGLGNLYDNWVKKYOMKNGSVFNSP 240
181 LSKOTDPSMLHKELEOKRCHSNEMDGYLAYSGLGNLYDNWVKKYOMKNGSVFNSP 240
241 SATAAAFINQHPGCLNLYNSLDKFCGNVPTVYPHDLFIRLSMVDTIERLGISSHFRVE 300
241 SATAAAFINQHPGCLNLYNSLDKFCGNVPTVYPHDLFIRLSMVDTIERLGISSHFRVE 300
301 IKNVLDTRCWRVERDEQIFMDVVTCTALAFRLRIRNGEYVSPDLAEITNELLAKDEYAA 360
301 IKNVLDTRCWRVERDEQIFMDVVTCTALAFRLRIRNGEYVSPDLAEITNELLAKDEYAA 360
361 LETYHSHILYQEDLSGKQILKASADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
361 LETYHSHILYQEDLSGKQILKASADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
421 RINTRENIQLYNVNDTRILKTTYHSSNISNTDYLRLAIVEDFYTCQSIYREELKGLERWV 480
421 RINTRENIQLYNVNDTRILKTTYHSSNISNTDYLRLAIVEDFYTCQSIYREELKGLERWV 480
481 ENKLDQKFARQKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDDFDGGTIDEL 540
481 ENKLDQKFARQKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDDFDGGTIDEL 540
541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSVHIQTWLEL 600
541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSVHIQTWLEL 600
601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVESSEYHNL 660
601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVESSEYHNL 660
661 FKLMSQGRLLNDIHSFKREFKQKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKREKEL 720
661 FKLMSQGRLLNDIHSFKREFKQKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKREKEL 720
721 MKLIFEENGSIIVPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
721 MKLIFEENGSIIVPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
781 EQOR 784
781 EQOR 784

RESULT 2
US-10-041-018-384
; Sequence 384, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080USI/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 384
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Stevia rebaudiana
US-10-041-018-384

Query Match 97.9%; Score 4026; DB 15; Length 784;
Best Local Similarity 98.0%; Pred. No. 2.7e-291;
Matches 768; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 MNLISCTASPLLTSSRPTALSAHTASTSHGGTNPNTNLIIDTTKRIOKLFKNVSI 60
Db 1 MNLISCTASPLLTSSRPTALSAHTASTSHGGTNPNTNLIIDTTKRIOKLFKNVSI 60

QY 61 SSYDTAWVAMPSPNPKSPCFPECLNWLINQNDGSLVNHHTNHNHPLKDSLSST 120
Db 61 SSYDTAWVAMPSPNPKSPCFPECLNWLINQNDGSLVNHHTNHNHPLKDSLSST 120
121 LACTIVALKRWNVGEDQINKGLSFIESNLASATEKSPSPIGFDIIPFGLLEYAKNLDINL 180
121 LACTIVALKRWNVGEDQINKGLSFIESNLASATEKSPSPIGFDIIPFGLLEYAKNLDINL 180
181 LSKOTDPSMLHKELEOKRCHSNEMDGYLAYSGLGNLYDNWVKKYOMKNGSVFNSP 240
181 LSKOTDPSMLHKELEOKRCHSNEMDGYLAYSGLGNLYDNWVKKYOMKNGSVFNSP 240
241 SATAAAFINQHPGCLNLYNSLDKFCGNVPTVYPHDLFIRLSMVDTIERLGISSHFRVE 300
241 SATAAAFINQHPGCLNLYNSLDKFCGNVPTVYPHDLFIRLSMVDTIERLGISSHFRVE 300
301 IKNVLDTRCWRVERDEQIFMDVVTCTALAFRLRIRNGEYVSPDLAEITNELLAKDEYAA 360
301 IKNVLDTRCWRVERDEQIFMDVVTCTALAFRLRIRNGEYVSPDLAEITNELLAKDEYAA 360
361 LETYHSHILYQEDLSGKQILKASADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
361 LETYHSHILYQEDLSGKQILKASADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
421 RINTRENIQLYNVNDTRILKTTYHSSNISNTDYLRLAIVEDFYTCQSIYREELKGLERWV 480
421 RINTRENIQLYNVNDTRILKTTYHSSNISNTDYLRLAIVEDFYTCQSIYREELKGLERWV 480
481 ENKLDQKFARQKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDDFDGGTIDEL 540
481 ENKLDQKFARQKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDDFDGGTIDEL 540
541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSVHIQTWLEL 600
541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSVHIQTWLEL 600
601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVESSEYHNL 660
601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVESSEYHNL 660
661 FKLMSQGRLLNDIHSFKREFKQKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKREKEL 720
661 FKLMSQGRLLNDIHSFKREFKQKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKREKEL 720
721 MKLIFEENGSIIVPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
721 MKLIFEENGSIIVPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
781 EQOR 784
781 EQOR 784

RESULT 3
US-10-041-018-390
; Sequence 390, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080USI/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 390
; LENGTH: 788
; TYPE: PRT
; ORGANISM: Lactuca sativa

US-10-041-018-395
; Sequence 395, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 395
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Cucurbita maxima
US-10-041-018-395

Query Match 49.2%; Score 2023.5; DB 15; Length 789;
Best Local Similarity 50.0%; Pred. No. 6.2e-142;
Matches 393; Conservative 150; Mismatches 212; Indels 31; Gaps 8;

QY 16 NRPAALGAIHTASTSHGQTNPTNLIIIDT-----TKERIQKQFNVEISVSYD 64
DB 4 SRPTGVARFAASSSSSSSASLFFGVDVDTTKTTCALHFBETKERIKKLDKVELSVSYD 63

QY 65 TAWVAVPSPNSPKSPCFPECLNWLINNQNDGSGWGLVNHNNHPLI-KDLSSTLAC 123
DB 64 TAWVAVPSPNSLNQPLFPECINVLVDSHADGSGWGL-----HNDQLMKANLLSTLAC 118

QY 124 IVALKRWNVGSDQNKGLSFIESNLASATKSOPSPIGRDIIPFGLLEYAKNLDINLSK 183
DB 119 VLTUKRWNIIGHDKWSKALDFIKSNIAATDENQSPVGFIIIPGMIEYAKNLDINLPLA 178

QY 184 QTFDSLMLHKRELBQKCHSNEMDG---YLAYISEGLNLYDNMVKYQKMGSGVNSPS 241
DB 179 PTNVDALVRKKELELRSCRSNSEGCKAYLAYVSEGGIKLQDWDVMVQYQKNGSLFNSPS 238

QY 242 ATAAAFINHQNPGLNVLNLLDKFGNAVTVVPHDLFIELSMVDVTERLIGTSHHPRVET 301
DB 239 TTAFAAFHRNDGCFDYLRSLLQKPDGSGVPTIPLDIYARLHWVDSLQKPGIARHPKEI 298

QY 302 KNVLDEYRCWVERDEQIFMDVVTICALAFRLRLINGEVSPPDLAETINEL-----AL 354
DB 299 RSVLDEYRCWQGEENIFLDASTCAMAFLRVBQVDSQDQIQPSEDIFFCGLGYL 358

QY 355 KDEVALETYHSHILYQEDLSSGKQILK-SADFLKEIISTD---SNRLSKLIHKEVENA 410
DB 359 KDFGASLELYRASQIITHPDSVLINENSWTSRFLKHLGSSDSVMSDRDTSVYKQAVNA 418

QY 411 LKFPINTGLERINRRNIQLNVNDRILKTYHSSNISNTDYLRLAVEDFYTCQIYRE 470
DB 419 LEFPNATLERLISKRAMESISGDIVISKSPYACINFGQDFLELAVEDFNTLQRIHLK 478

QY 471 ELKGLERWVENKLDQLKFAKQKTAYCYFSVAATLSSPELSDARISWAKNGILTTVVDDF 530
DB 479 ELEELQRWVENKLDLKKFFRLHLGYCYFAAAATLTDPELHDARIAWAQGVLTITVVDDF 538

QY 531 FDIJGTIDELNLIQCYEKNVVDKCCSEHVRILFLALKDAILCWDGEAFKXQARDVT 590
DB 539 YDGGSEELONLIELVEKWDPDGEGVCKDVEIFVALHSTVCEIGRRALVWQGSVM 598

QY 591 SHVITWLELMNSMLREAITRDVAVPTPLNEYMENAVSPFALGPVVKPAIFYGPKLSEE 650
DB 599 RNVIDGWLALLKVMKEAEWSTNVKPSMGEBYQAHVSPALGPVILPLMFFVGPKLSEE 658

QY 651 IVESSEYVNLKLMSTQRLANDIHSFKREFKEKINAVLHLSNGESKVEEVEEVM 710
DB 659 MIGSEYOKLYKMSIAGRLKNDIRSYDRECKEKLNLISLWMDGGGNVTKEEAIEAK 718

QY 711 MMKNKRKELMKLIPFEENGSIIVPRACKDAFNMWCHVNFYFANDDGGFTNTILDTVKDII 770

QY 530 FFDIGTIDETNLITQVEKKNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWADV 589
DB 522 FFDVGSKEELENLALVEKHAHVAFFYSQVXIVSAIYTTVNHGASAAQGRDL 581
QY 590 TSHVQTWLELWNSMLRAIWRDRAVPTLNHMYENAVSFALGPVIVKPAIFVFGPKLSE 649
DB 582 TNHLVEIWLDIRSNMVAEAWQRCQVPTVEYMTNAVVSFALGPVILPALYFVQELLE 641
QY 650 EIVSESHNLPKLMSTGRLLNDLHSPKEPKGLNAVALHLSNGSGKVEEVEEM 709
DB 642 HAVKDEEYKDFRLVSTGRLLNDVQSLEREGNOGKLSVSLVLHSGGSMSEIAAKAM 701
QY 710 MAMINKKELMKLIFEENGSIIVPRACKDAFNMCHVILNFFYANDDGTG-NTIIDLTKD 768
DB 702 QKSIDVSRDLRLVLRKE-SAVPRCKELFWCKILHLFLFYSQNDGFSSPKEMYSAYNA 760
QY 769 ILYNPLVLVN 778
DB 761 VINBPLKVQN 770

RESULT 7

US-10-259-194A-108
; Sequence 108, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghasseman, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Krepis, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (c) 2001 Syngenta
; SEQ ID NO 108
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-194A-108

Query Match 37.4%; Score 1537.5; DB 15; Length 740;
Best Local Similarity 41.7%; Pred. No. 1e-105;
Matches 325; Conservative 142; Mismatches 243; Indels 69; Gaps 16;
QY 20 ALSAHTASTSHGGQTNTNLIIDTTKERIOKFNVEISVSYDTANVAVPS-PNSPK 78
DB 5 AMRHCSSSSEEGAAATTAARSAREL--QLAPSPSPYDTANVAVPALRRGGG 62
QY 79 SPCPECLNWLNNQDGSW--GLVNHNNHP--LLKSLSTLACTIVALKRWNVGE 134
DB 63 GPRFPQCVAMIQNRQGDGWRHAAAHQQLGSSPEIVTERDLSLSTLACVLARWDAGS 122
QY 135 DQINKGLSFISNLASATEKQSPPIGFDII-FPGLLLEYAKNLDINLIS-----KQTFD 187
DB 123 EHVRRGLQFIGRNMVAMDDQTAAPASGVSVFAAMLRAMEMWGLEVPASQADVRDRDA 182
QY 188 SIMLHKRELEQKRCHSNEMDGYLAYSISGLNLYDWNVVKYQMKNGSVFNSPSATAAF 247

DB 183 GVI-----CHGGRTY-YTAVSGLGNIQWNEVMKFORNGSLFNSPYTAAAL 231
QY 248 INHONPGCLNLYNSLLDKFGNAVPTVYPHDLFIRLSMVDTIERLGI SHHFVRKVLDE 307
DB 232 VHYDAKALQYLDMLDKFGSAVPAAYPANIQSOLYMDVLEKMGISRHFVGEIKSILDM 291
QY 308 TYRCWVERDEOIFMDVVTICALAFRLRLRINGVEVSDPDLAEIT-----NELA--LKDEYA 359
DB 292 TYSCKWQDEEIVLDMQTCGAFRLMRNGVDVSDLSHFSEPSFNSLQGYLNDTRS 351
QY 360 ALETYHSHILYQEDLSSGKQILKSADFLKETIISTDSNRLSKLIHKEVENALKFPINTGL 419
DB 352 LLELHKASKV-----SIAEKEVEYALEFFPYTIL 380
QY 420 BRINTRMIQYVNDVNTIRLKTTHSSNISNDYLRVLAVDFYTQCSYRRELKGLERWV 479
DB 381 DRLDHRKRIEHFDITSSQMLETAVLPCH-SNEEIMALGVDRDFSSQFIFQBELQQLNSWV 439
QY 480 VENKLDQLKEAROKTAYCYFSVAATLSSPELSDARISWAKNGILTTVVDDDFDIDGTTDE 539
DB 440 KESRUDQOFARQKLDYFYFSAATITFPELSDVRILWAKNGVLTVVDDDFDVGSGKEE 499
QY 540 LTNLITQVEKKNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWADVTSHTVITWLE 599
DB 500 LENLVALVEKNDKTEYYSQVEIVFSAIYTSNQLGSMASVVOGRDVTKHLVEIWE 559
QY 600 LMNSMLREAIWTRDAYVPTLNHMYENAVSFALGPVIVKPAIFVFGPKLSEIEVESSEYHN 659
DB 560 LRSMTTEVWROSRVYPTABEYMENAVVTFALGPVILPALYFVQPKIPDSVIRSQCSE 619
QY 660 LFKLMSTGRLLNDIHSFKREFKCKLNAVALHLSNGSGKVE-EEVVEEMMMIKNRK 718
DB 620 LFRMSKGRLLNDVQSYEREGSQGLNSVSL-LALHSGGVSVEEAVKQIQRPLEKCR 678
QY 719 ELMKLIFFEENGSIIVPRACKDAFNMCHVILNFFYANDDGTGNTIIDLTKD-IYINPLVL 776
DB 679 ELLKLVSRGGA-VPRPCRELFWNSCKVCHFFYSGDGFSSPTAKAGALDAVIHEPLNL 736

RESULT 8

US-10-437-963-121564
; Sequence 121564, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121564
; LENGTH: 689
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(689)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24577C.1.pap
US-10-437-963-121564

Query Match 35.4%; Score 1458; DB 16; Length 689;
Best Local Similarity 41.9%; Pred. No. 7.7e-100;

```

RESULT 9
US-10-437-963-138526
; Sequence 138526, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

```

RESULT 10
US-10-437-963-174639
; Sequence 174639, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

QY 456 LAVEDPYTQSIYREELKGLERWVENKLDQLKFAKQKTAICYFSSVAATLSSPELSDARI 515
Db 474 LAKLDFNMVQALHOKETOHIWVSWRESGENDLFTTRQRPVEMVFSVAVSMFEPEFAACRI 533
QY 516 SWAKNGILTVVDDPFDIGTIDELTNLQCVKKNVDVDDCCSEHVRILFLALKAIDC 575
Db 534 AYAKTSCLAVILDDLYTHGSLDDELKLFSEAVRMDISVLDSVRDQKLVCFGLGLNTVN 593
QY 576 WIGDEAFKQWQARDVTSHTVQTLWELMNSMLREAIWTRDAYVPTLNEYMENAVYSPALGPI 635
Db 594 GFQKGLKEQGRDVLGYLAKVWEGLLASTYKAEWSAAKYVPTFNEYENAKVSIATV 653
QY 636 VKPAIVFGPKLSEIVESEYHNLF-KLMSTQGRLLNDIHSFKRFEKGL-NAVALHL 693
Db 654 VLNSIFFTGELLPDYILQOVDLRSKFLHLVSLTGRLLNDTKTYQAEERNRGELVSSVQCYM 713
QY 694 -SNGESKVEEVEEMMMIKKRELKMLKLFENGSIIVPRACKDAFNMCHVLNFFYA 752
Db 714 RENPEC--TEEEALSHVYGIIDNALKEL-NWELANPASNAPLCVRRLLENTARVMQLFTYM 770
QY 753 NDDGF--TCNTILDTVKDIIYNPL 774
Db 771 YRDGFGISKEMKDHVSRLFDPV 794
RESULT 14
US-10-041-007-35
; Sequence 35, Application US/10041007
; Publication No. US20020164736A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Schepmann, Hala G
; TITLE OF INVENTION: Ginkgo Biloba Levopimaradiene Synthase
; FILE REFERENCE: P02081US1
; CURRENT APPLICATION NUMBER: US/10/041, 007
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259, 881
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 814
; TYPE: PRT
; ORGANISM: Ginkgo biloba
US-10-041-007-35
Query Match 26.1%; Score 1075; DB 13; Length 814;
Best Local Similarity 32.3%; Pred. No. 3.6e-71;
Matches 260; Conservative 159; Mismatches 329; Indels 56; Gaps 16;
QY 10 PLLTKSNRPAALSALHTASTSHGGQTNPTNLIIDTTKERIQOKNV---EISVSSYDTA 66
Db 27 PAVWKDDFIDSLTSPNSHATSKSSVDETKIKETLVKEIQCMFQSMGSGDGTNPAYDTA 86
QY 67 WVAWVPSNPKSPCFPECLNWLINQNDGSGVLNHNHPLKDLKSLSTLACTIVA 126
Db 87 WVARIPSDGSGAPQFPQTLQWILNNQLPDGSGW-----EBCIFLAYDRVLNTLACLT 140
QY 127 LKRWNVGDOINKGLSFIESNLAS-ATEKSPSPIGFDIIPGLLEYAKNLDINLLSKQT 185
Db 141 LKIWNKGDIQVKQGVFFVRKHEEMKDEADNHRPSGFVFPFAMLDLGLDLPYHLP 200
QY 186 DFLMLHKLRELEQKRECHENMDGY---LAYISEGLNLYDNWNVKVKYQKNGSVFNSPSA 242
Db 201 FISQIHQRQKQKLOKIPLVNLHNHQTALLYSLGLEQDVVDVQBEITNLQROGSGFSSPAS 260
QY 243 TAAAPINONPCGINYLNSLDKFGNAVTVYVPHDLFIRLSMVDTIERLIGISHFRVEIK 302
Db 261 TACVPMHTQNRKCLHFLNLSKFGDYVCPFLDLFERLWAVDIVERLIGIDRYFKKIK 320
QY 303 NVLDSEYECW-VERDE-----QIFMDVVTALAFRLIRINGEVSPDPFAEITNE----- 351
Db 321 ESLDYVYWDARGVGWARCNPIPDVDDTAMGLRILRLHGVNVSDDVLENFRDEKGDFF 380

QY 352 -----LAKDEYAALETYHASHILYQEDLSGKQILKSA-----DFLKKEIISTD 395
Db 381 CFAGOTQIGVTN---LNLRCRSQVCF-----PGEKIMEAKTFTTNHLQNALAKNNAFD 432
QY 396 SNRLSKLIHKEVENALKPPIINTGLERINTRRMIOLYNVNDTRILTKTYHSSNISNTDYL 455
Db 433 KWAVKDLFGVEYAIKYPWHRSMFLEARSYIEQPGSNDVWLGKTVYKMLYVSNKYLE 492
QY 456 LAVEDPYTQSIYREELKGLERWVENKLDQLKFAKQKTAICYFSSVAATLSSPELSDARI 515
Db 493 LAKLDFNMVQALHOKETOHIWVSWRESGENDLFTTRQRPVEMVFSVAVSMFEPEFAACRI 552
QY 516 SWAKNGILTVVDDPFDIGTIDELTNLQCVKKNVDVDDCCSEHVRILFLALKAIDC 575
Db 553 AYAKTSCLAVILDDLYTHGSLDDELKLFSEAVRMDISVLDSVRDQKLVCFGLGLNTVN 612
QY 576 WIGDEAFKQWQARDVTSHTVQTLWELMNSMLREAIWTRDAYVPTLNEYMENAVYSPALGPI 635
Db 613 GFQKGLKEQGRDVLGYLAKVWEGLLASTYKAEWSAAKYVPTFNEYENAKVSIATV 672
QY 636 VKPAIVFGPKLSEIVESEYHNLF-KLMSTQGRLLNDIHSFKRFEKGL-NAVALHL 693
Db 673 VLNSIFFTGELLPDYILQOVDLRSKFLHLVSLTGRLLNDTKTYQAEERNRGELVSSVQCYM 732
QY 694 -SNGESKVEEVEEMMMIKKRELKMLKLFENGSIIVPRACKDAFNMCHVLNFFYA 752
Db 733 RENPEC--TEEEALSHVYGIIDNALKEL-NWELANPASNAPLCVRRLLENTARVMQLFTYM 789
QY 753 NDDGF--TCNTILDTVKDIIYNPL 774
Db 790 YRDGFGISKEMKDHVSRLFDPV 813
RESULT 15
US-10-041-007-2
; Sequence 2, Application US/10041007
; Publication No. US20020164736A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Schepmann, Hala G
; TITLE OF INVENTION: Ginkgo Biloba Levopimaradiene Synthase
; FILE REFERENCE: P02081US1
; CURRENT APPLICATION NUMBER: US/10/041, 007
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259, 881
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Ginkgo biloba
US-10-041-007-2
Query Match 26.1%; Score 1075; DB 13; Length 873;
Best Local Similarity 32.3%; Pred. No. 4e-71;
Matches 260; Conservative 159; Mismatches 329; Indels 56; Gaps 16;
QY 10 PLLTKSNRPAALSALHTASTSHGGQTNPTNLIIDTTKERIQOKNV---EISVSSYDTA 66
Db 86 PAVWKDDFIDSLTSPNSHATSKSSVDETKIKETLVKEIQCMFQSMGSGDGTNPAYDTA 145
QY 67 WVAWVPSNPKSPCFPECLNWLINQNDGSGVLNHNHPLKDLKSLSTLACTIVA 126
Db 146 WVARIPSDGSGAPQFPQTLQWILNNQLPDGSGW-----EBCIFLAYDRVLNTLACLT 199
QY 127 LKRWNVGDOINKGLSFIESNLAS-ATEKSPSPIGFDIIPGLLEYAKNLDINLLSKQT 185
Db 200 LKIWNKGDIQVKQGVFFVRKHEEMKDEADNHRPSGFVFPFAMLDLGLDLPYHLP 259
QY 186 DFLMLHKLRELEQKRECHENMDGY---LAYISEGLNLYDNWNVKVKYQKNGSVFNSPSA 242

Db 260 FISQIHQKQKQLQKIPILNVLHNOTALLYSLGLQDVVDWQBIITNLSQSDGSLSSPAS 319
QY 243 TAAAFINONPGCLNVLNLLDKFGNAVTVPHDLPIRLSMVDTIERLIGISHHPRVEIK 302
Db 320 TACVFMHTQKRCLEHFNFLSKFGDYVPCYPLDLFERLWAVDVTVLIGIDRYFKKEIK 379
QY 303 NVLDEYRCW-VERDE---QIFMDVVTICALAFRLRLRINGEVSPPDLAEITNE----- 351
Db 380 ESLDYVYRWDAERGCVGWARCNIPDPVDDTAMGLRLRLHGNVSSDVLENFRDEKGDFF 439
QY 352 -----LALKDEYAALEYHSHILYQEDLSSGKOILKSA-----DFLKEIISTD 395
Db 440 CFAGQOIGVTDN---LNLYRCSQVCF---PGEKIMEEAKFTTNNHLQNALAKNNAFD 491
QY 396 SNRLSLIHKVENALKFPINTGLERINTFRNIQLYNVDNTRILKTYHSSNISNTDYL 455
Db 492 KNAVKKDLPEGEVEYAIKYPWHRSMPLREARSYIEQFGSNDVWLKGTVYQXLYVSNKYLE 551
QY 456 LAVEDFVTCQSIYREELKGLERWVENKLDQKFARQKTAYCYFSVAATLSSPELSDARI 515
Db 552 LAKLDFNMVQALHOKETOHIVSWRESGFNDLFTFQRPVENYFSVAVSMFEPEFAACRI 611
QY 516 SWAKNGILFTVVDDDFDIDGTIDELNLIQCEKKNVVDKCCSHVRILFLALKDAIC 575
Db 612 AVAKTSCLAVILDDLYDTHGSLDDLKLFSEAVRRWDISVLDSVRDNQLKVCFTGLYNTVN 671
QY 576 WIGDEAFKQARDVTSHVICTWLELNNMLREAIWTDAYVPTLNEYMENAYVSEALGPI 635
Db 672 GFKGGLKEQGRDVLGYLRKRWEGLLASYTKEAENSAAKYVPTFNEYVENAKVSIALATV 731
QY 636 VKPAIYFVGPKLEEIVSESEYHNLF-KLMSTOGLINDIHSFKRBFKEGKL-NAVALHL 693
Db 732 VLNSIFTGELLDPYILQQVDLRSKFHLVSLTGLRLINDTKTYQAEENRGELVSSVQCYM 791
QY 694 -SNGESGKVEEVEEENMMIKRKRKELMKLIFEENGSIIVPRACKDAFWMCHVLNFFYA 752
Db 792 RENPEC--TEEEALSHVYGIIDNALKEL-NWELANPASNAFLCVRRLLENTARVMQLFTYM 848
QY 753 NDDGF--TGNITLDYKDIYNPL 774
Db 849 YRDGFGISDKEMKDHVSRTLFDPV 872

Search completed: October 17, 2004, 09:18:45
Job time : 75.5657 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:24:39 ; Search time 210.188 Seconds
(without alignments)
1338.060 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSICIASPLTKSNRPAA.....TVKDIYNPLVNVNEBQR 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 4113 | 100.0 | 784 | 8 | Adm98963 Diterpene |
| 2 | 4026 | 97.9 | 784 | 8 | Adm98964 Diterpene |
| 3 | 2855.5 | 69.4 | 788 | 8 | Adm98970 Diterpene |
| 4 | 2101.5 | 51.1 | 785 | 8 | Adm98969 Diterpene |
| 5 | 2023.5 | 49.2 | 789 | 8 | Adm98975 Diterpene |
| 6 | 2021 | 49.1 | 785 | 5 | Adm98975 Diterpene |
| 7 | 2012.5 | 48.9 | 789 | 2 | Adm98975 Diterpene |
| 8 | 1537.5 | 37.4 | 740 | 8 | Adm98975 Diterpene |
| 9 | 1523 | 37.0 | 732 | 8 | Adm98975 Diterpene |
| 10 | 1075 | 26.1 | 873 | 6 | Adm98975 Diterpene |
| 11 | 1075 | 26.1 | 873 | 8 | Adm98975 Diterpene |
| 12 | 1062.5 | 25.8 | 868 | 3 | Adm98975 Diterpene |
| 13 | 1062.5 | 25.8 | 868 | 8 | Adm98975 Diterpene |
| 14 | 1062.5 | 25.8 | 868 | 6 | Adm98975 Diterpene |
| 15 | 1051.5 | 25.6 | 868 | 6 | Adm98975 Diterpene |
| 16 | 1015.5 | 24.7 | 862 | 6 | Adm98975 Diterpene |
| 17 | 1013.5 | 24.6 | 862 | 2 | Adm98975 Diterpene |
| 18 | 1013.5 | 24.6 | 862 | 3 | Adm98975 Diterpene |
| 19 | 1013.5 | 24.6 | 862 | 5 | Adm98975 Diterpene |
| 20 | 1013.5 | 24.6 | 862 | 8 | Adm98975 Diterpene |
| 21 | 965 | 23.5 | 782 | 3 | Adm98975 Diterpene |
| 22 | 965 | 23.5 | 782 | 3 | Adm98975 Diterpene |
| 23 | 965 | 23.5 | 782 | 4 | Adm98975 Diterpene |
| 24 | 965 | 23.5 | 782 | 5 | Adm98975 Diterpene |
| 25 | 965 | 23.5 | 817 | 2 | Adm98975 Diterpene |

| | | | | | |
|----|-------|------|-----|---|--------------------|
| 26 | 965 | 23.5 | 817 | 2 | AAY06566 Grand fir |
| 27 | 965 | 23.5 | 817 | 2 | AAY06562 Grand fir |
| 28 | 965 | 23.5 | 817 | 5 | Abb79391 Taxadiene |
| 29 | 965 | 23.5 | 817 | 6 | Abu09781 White fir |
| 30 | 962 | 23.4 | 817 | 2 | AAY06568 E-alpha-b |
| 31 | 925 | 22.5 | 458 | 8 | Adm98972 Diterpene |
| 32 | 863 | 21.0 | 870 | 3 | AAY70719 Clarkia b |
| 33 | 863 | 21.0 | 870 | 6 | AAY70719 Clarkia b |
| 34 | 854.5 | 20.8 | 884 | 5 | AAB91405 Herbicide |
| 35 | 823 | 20.0 | 870 | 2 | AAW17080 Clarkia b |
| 36 | 823 | 20.0 | 870 | 2 | AAW17080 Clarkia b |
| 37 | 752.5 | 18.3 | 811 | 8 | Adm94222 Soybean E |
| 38 | 743 | 18.1 | 815 | 3 | AAY84104 Soybean c |
| 39 | 743 | 18.1 | 815 | 5 | AAY84104 Soybean c |
| 40 | 732 | 17.8 | 787 | 8 | Adm98965 Diterpene |
| 41 | 725.5 | 17.6 | 801 | 8 | Adm98974 Diterpene |
| 42 | 709 | 17.2 | 802 | 5 | AAW90848 Gibberell |
| 43 | 709 | 17.2 | 802 | 5 | AAW90848 Gibberell |
| 44 | 657.5 | 16.0 | 823 | 8 | Adm98976 Diterpene |
| 45 | 656 | 15.9 | 823 | 2 | AAW93800 Anthr ea |

ALIGNMENTS

RESULT 1

ADM98963
ID ADM98963 standard; protein; 784 AA.

XX AC ADM98963;

XX DT 01-JUL-2004 (first entry)

XX DE Diterpene synthase polypeptide #1.

XX KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
XX KW diterpene synthase; defence toxin; volatile defensive signal;
XX KW pollinator attractant; photoprotectant; enzyme.

XX OS Stevia rebaudiana.

XX PN US2004072323-A1.

XX PD 15-APR-2004.

XX PF 07-JAN-2002; 2002US-00041018.

XX PR 05-JAN-2001; 2001US-0259880P.

XX PA (MATS/) MATSUDA S P T.

XX PA (HART/) HART E A.

XX FI Matsuda SPT, Hart EA;

XX DR WPI; 2004-373921/35.

XX PT New unicellular organisms comprising exogenous nucleic acids encoding a
XX PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX PT producing diterpenes and diterpene precursors.

XX PS Claim 31; SEQ ID NO 383; 38pp; English.

XX CC The invention relates to a unicellular organism for producing a diterpene
XX CC or diterpene precursor comprising an exogenous nucleic acid sequence
XX CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
XX CC promoter operable in the organism, and an exogenous nucleic acid sequence
XX CC encoding a diterpene synthase under the control of a promoter operable in
XX CC the organism. The invention also relates to methods of producing a
XX CC diterpene or diterpene precursor and a method of isolating a diterpene
XX CC synthase comprising growing several cells in the presence of a
XX CC polyaromatic resin to make a cell/resin mixture, where at least one of
XX CC the cells further comprises at least one isolated and purified nucleic
XX CC acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.

| | | |
|----|---------------------------|---|
| XX | Sequence 784 AA; | |
| SQ | Query Match | 100.0%; Score 4113; DB 8; Length 784; |
| | Best Local Similarity | 100.0%; Pred. No. 0; |
| | Matches 784; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; |
| QY | 1 | MNLSLCIASPLITKSNRPAALSAIHTASTSHGGQTNTNLIIIDTTKRIQKQFNVEISV 60 |
| DB | 1 | MNLSLCIASPLITKSNRPAALSAIHTASTSHGGQTNTNLIIIDTTKRIQKQFNVEISV 60 |
| QY | 61 | SSVDTAWAVSPNSPKPCPECLNLIINLNQDGSWGLVNHHTNHHPILLKDSLSST 120 |
| DB | 61 | SSVDTAWAVSPNSPKPCPECLNLIINLNQDGSWGLVNHHTNHHPILLKDSLSST 120 |
| QY | 121 | LACIVALKRWNVGEDQINKGLSPFIESNLASATEKSPSPIGFDIIIPFGLLEYAKNLDINL 180 |
| DB | 121 | LACIVALKRWNVGEDQINKGLSPFIESNLASATEKSPSPIGFDIIIPFGLLEYAKNLDINL 180 |
| QY | 181 | LSKQTFSLMLHRELEOKRCHENMDGYLAIYISEGLGNLYDNWNVKKYQMKNGSVFNSP 240 |
| DB | 181 | LSKQTFSLMLHRELEOKRCHENMDGYLAIYISEGLGNLYDNWNVKKYQMKNGSVFNSP 240 |
| QY | 241 | SATAAFINHQPCCLNLYNSLADKCGNAVPTVYPHDLFRLSWMDTIERLGISSHFRVE 300 |
| DB | 241 | SATAAFINHQPCCLNLYNSLADKCGNAVPTVYPHDLFRLSWMDTIERLGISSHFRVE 300 |
| QY | 301 | IKNVLDIETRCWVERDEQIFMDVVTCALAFRLRINGEYVSPDPLAIBITNELAKDEYAA 360 |
| DB | 301 | IKNVLDIETRCWVERDEQIFMDVVTCALAFRLRINGEYVSPDPLAIBITNELAKDEYAA 360 |
| QY | 361 | LETYHSHILYQEDLSGKQILKSADFLKEIISTDSNRLSKLIHKEVENALKFPINTGLE 420 |
| DB | 361 | LETYHSHILYQEDLSGKQILKSADFLKEIISTDSNRLSKLIHKEVENALKFPINTGLE 420 |
| QY | 421 | RIMTRNIOQYVNDNTRILKTTVHSSNINSNTDYLRLAIVEDFYTCQSIYREELKGLERWV 480 |
| DB | 421 | RIMTRNIOQYVNDNTRILKTTVHSSNINSNTDYLRLAIVEDFYTCQSIYREELKGLERWV 480 |
| QY | 481 | ENKLDQKFAKQRTAYCFYVAATLSSPELSDARISWAKNGILTTVVDDPFDIGGTIDEL 540 |
| DB | 481 | ENKLDQKFAKQRTAYCFYVAATLSSPELSDARISWAKNGILTTVVDDPFDIGGTIDEL 540 |
| QY | 541 | TNLIQCKEKNVDYDKCCSEHVRILPLAKDAICWIGDFAKWAQADVTSHVITQWLEL 600 |
| DB | 541 | TNLIQCKEKNVDYDKCCSEHVRILPLAKDAICWIGDFAKWAQADVTSHVITQWLEL 600 |
| QY | 601 | MNSMLREAIWTRDAYVPTLNEYMENAVSPALGPVFKPAIFVFGPKLSEIVSESYHNL 660 |
| DB | 601 | MNSMLREAIWTRDAYVPTLNEYMENAVSPALGPVFKPAIFVFGPKLSEIVSESYHNL 660 |
| QY | 661 | FKLMSTQGRLLNDIHSFKRFBKEGKLNALVALHLSNGSGKVEEBVVEEMMMIKNKKREL 720 |
| DB | 661 | FKLMSTQGRLLNDIHSFKRFBKEGKLNALVALHLSNGSGKVEEBVVEEMMMIKNKKREL 720 |
| QY | 721 | MKLIIFENGSIIVPRACKDAFNMWCHVLNFFVANDDGTGTILDTVKDIIYNPLVLNEN 780 |
| DB | 721 | MKLIIFENGSIIVPRACKDAFNMWCHVLNFFVANDDGTGTILDTVKDIIYNPLVLNEN 780 |
| QY | 781 | EEQR 784 |

| | | |
|----------|---|----------------------------|
| DB | 781 | EEQR 784 |
| RESULT 2 | | |
| ADM98964 | | |
| ID | ADM98964 | standard; protein; 784 AA. |
| XX | ADM98964; | |
| XX | 01-JUL-2004 | (first entry) |
| XX | Diterpene synthase polypeptide #2. | |
| XX | Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor; | |
| XX | diterpene synthase; defence toxin; volatile defensive signal; | |
| XX | pollinator attractant; photoprotectant; enzyme. | |
| XX | Stevia rebaudiana. | |
| XX | US2004072323-A1. | |
| XX | 15-APR-2004. | |
| XX | 07-JAN-2002; | 2002US-00041018. |
| XX | 05-JAN-2001; | 2001US-0259880P. |
| XX | (MATS/) MATSUDA S P T. | |
| XX | (HART/) HART E A. | |
| XX | Matsuda SPT, Hart EA; | |
| XX | WPI; 2004-373921/35. | |
| XX | New unicellular organisms comprising exogenous nucleic acids encoding a | |
| XX | geranylgeranyl pyrophosphate and a diterpene synthase, useful for | |
| XX | producing diterpenes and diterpene precursors. | |
| XX | Claim 31; SEQ ID NO 384; 38pp; English. | |

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.

| | | |
|----|---------------------------|---|
| XX | Sequence 784 AA; | |
| SQ | Query Match | 97.9%; Score 4026; DB 8; Length 784; |
| | Best Local Similarity | 98.0%; Pred. No. 0; |
| | Matches 768; Conservative | 8; Mismatches 8; Indels 0; Gaps 0; |
| QY | 1 | MNLSLCIASPLITKSNRPAALSAIHTASTSHGGQTNTNLIIIDTTKRIQKQFNVEISV 60 |

Db 293 VDTLRLGIGKRFMVEIQNVLDITYRCVWQGDVQIFMDVVTICALAFRLVRSNGYEVSSDP 352
QY 345 LAETYN-----FLAKDEYAALETYHSHILYOEGLSSGKOLKSGADFLKEIISTDS 396
Db 353 LAKITKEGDMNSPEKPKDQVTSLEVYKASQIIYOEELAFREQNLT-----YLPSS 405
QY 397 NLSKLIHKEVENALKFPINTGLERINTRRNIQLYNVNTRILKTYTHSSNISNTDYLR 456
Db 406 NKLSNYILKEVDALKFPNGSLRMSRNEIHNHTRILKTYSSNISNKLK 465
QY 457 AVEDFYTCQSYREELKGLRWVVENKLDOLKFAKQTAICYFVVAATLSSPELSDARIS 516
Db 466 AVODFNEQOSIYCEELKOLRWVVENRLDKLFAKQTAICYFSAASFSSPDLSDARIS 525
QY 517 WAKNGILTVVDVDFDGTIDELTNLIQCVEKNVDVDDKCCSEHVRILFLAKDAICW 576
Db 526 WAKSSILTVTVDVDFDVGSDDELNVFVHIIIBKNVNVNENDCCSEVGVLFALAKDAICW 585
QY 577 IGDEAFKQARDVTSHVITQWLELMSMLREAIWTRDAYVPTLNEYMENAYVSFALGPV 636
Db 586 IGDKAFKIQERNITSHVIEIWDLVKSMLEALWAKDGSIPITNEYWENGYSFALGPV 645
QY 637 KPAIVFVGKLSSEIVSESYENLFLKMSQORLNDIHSFKREBFKGLNAVALHLSNG 696
Db 646 LPTLYFLGVKLSSEVVQSSEYHKLVEVMSQORLNDIHSFKREKAGKLNVALVYMSDG 705
QY 697 ESKGVEEVEEVMWMIKKNRKLAKLIPEENGSIYPRACKDAFNMCHVLNPFYANDDG 756
Db 706 KSGSVEEVEEVMKILTSQRKMMKLVLETKGSVVRVCKDVFNMCMVNLNLFYATDDG 765
QY 757 FTGNITLDVKDIIYNPL 774
Db 766 FTGNALDVVKELIIEPV 783

RESULT 4

ADM98969

ID ADM98969 standard; protein; 785 AA.

AC ADM98969;

XX

XX 01-JUL-2004 (first entry)

XX Diterpene synthase polypeptide #7.

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

XX diterpene synthase; defence toxin; volatile defensive signal;

XX pollinator attractant; photoprotectant; enzyme.

XX Cucumis sativus.

OS US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.

XX Sequence 785 AA;

Query Match 51.1%; Score 2101.5; DB 8; Length 785;
Best Local Similarity 50.9%; Pred. No. 1.7e-162;
Matches 413; Conservative 136; Mismatches 194; Indels 69; Gaps 10;

QY 1 MNLSLCIASPLTKSNRPALSAIHTASTSH-----GGQTNFTNLIIDTKERIQQ 52
Db 1 MNLS-----RPTNLGCF-TASSASLFFPLDVGTTKTGALFEETKEIRKKL 47

QY 53 FXNVELSVSYDTAVWAVVPNSPKSPCFPCPLANLNNOLNDSGLVNNHNNHPL 112
Db 48 FXNVELSAYDTAVWAVVPNSLNKPLFPCCINWLDQHPDQSGWGL-----HDHQL 102

QY 113 -LKDSLSSTLACTIVALKRWNVGEDIQINKGLSFIESNLASATEKSPSPGFIIFPGLLE 171
Db 103 VMKATLSTLACVLTWKWDIGDDHMSKALSIKSNIASATDENQSPVDFIIFPGMIE 162

QY 172 YAKNLDINLLSKOTDFSLMLHRELEQKCHENMDG--YLAISEGLNLDVWNNVKY 229
Db 163 YAKNLDINLPLASMNVDALVQKLELRSCSNSEGGKAYLAYVSEGIQKQDWENVMRY 222

QY 230 QMKGVSFNSPSATAAFINHQPCGLNVLNSLDKFGNAVPTVYPHDLFIRLSMVDTE 289
Db 223 QRRNGSLFSSPSTTAVAFMRNRDQDGCNFDVRSVLQKPHSVPAIYPLDIYARLHWDSLQ 282

QY 290 RLGISHHFRVEIKNVLDITYRCVWERDEQIFMDVVVTCALAFRLLRINGEVSPDPAEIT 349
Db 283 KLIGDGHFKDEIRSVLDITYSCWMQGEENIFLDASTCMAFMRVLEGYDVSSDQLTQFS 342

QY 350 NEL-----ALKDEYAALETYHSHILYOE-----LSSGKQLKS 384
Db 343 EGLFNSCLGHLKDFGSALELFAKASQIIYPPDFILENINSWTSRFINHLSGSG----- 396

QY 385 ADPLKEIITDSNRLSKLIHKEVENALKPPINTGLERINTRRNIQLYNVNTRILKTYTH 444
Db 397 -----SVHSDRTERLVKQEAUNAFAFPYNTSLERLSNKRALYESYSGDIVRISKAYA 448

QY 445 SSNISNTDYRLAIVEDFYTCQSYREELKGLRWVVENKLDOLKFAKQTAICYFVVAAT 504
Db 449 CLNFGHQDFLELAVEDFNTLQIHRKELKELEKWIENKLDKLKFAKQIAYCYFSAAT 508

QY 505 LSPFELSDARISWAKNGILTVVDVDFDGTIDELTNLIQCVEKNVDVDDKCCSEHVR 564
Db 509 LTSPELCDARLSWAKNGVLTTVVDVDFDVGSGSEELVNLILQVKEKWDASGETGYCSKEVE 568

QY 565 ILFLALKDAICWIGDEAFKQARDVTSHVITQWLELMSMLREAIWTRDAYVPTLNEYME 624
Db 569 IIFLALHSTICEIGKALKALPWQGSVVRNVIDIWLALLESNRKEAEWLKNKVVPSLDEYME 628

XX New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpenes and diterpene precursors.

XX Claim 31; SEQ ID NO 389; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene

XX Herbicidally active polypeptide SEQ ID NO 874.
 XX Herbicidal; plant; agriculture; herbicide.
 XX Arabidopsis thaliana.
 XX WO200210210-A2.
 XX 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX (PAB) BAYER AG.
 XX Tietjen K, Weidler M;
 XX WPI; 2002-269010/31.
 XX Identifying plant target proteins for herbicidally active compounds.
 XX comprising aligning and comparing nucleic acid or amino acid sequences
 XX from plant with nucleic acid or amino acid sequences from non-plant
 XX organisms.
 XX Claim 5; SEQ ID NO 874; 261pp + Sequence Listing; English.
 XX The invention relates to identifying target proteins (ABB90790-ABB94016)
 XX for herbicidally active compounds, comprising aligning and comparing
 XX nucleic acid or amino acid sequences from plant with nucleic acid or
 XX amino acid sequences from non-plant organisms using suitable search
 XX parameters, where plant sequences having an E-value greater by a factor
 XX of 3 than the E-value of most similar non-plant sequences are selected.
 XX The polypeptides or nucleic acids encoding them are useful for
 XX identifying modulators. The identified modulators are useful as
 XX herbicides
 XX Sequence 785 AA;
 XX
 Query Match 49.1%; Score 2021; DB 5; Length 785;
 Best Local Similarity 52.6%; Pred. No. 6.6e-156;
 Matches 402; Conservative 130; Mismatches 208; Indels 24; Gaps 9;
 QY 34 QTNPTLIIIDTKERIQKQKVEISVSYDTAWAVPSPNSPKPCFPECCLNLIINNO 93
 DB 27 QTRANNVFEQTEKIKMLKVELSVAYDSWAVPSPSQMAPLPFQCVKLLDNQ 86
 QY 94 LNDGSLVNHHTHHNHPLLKDSLSLTLACIVALKEMNVGEGDQINKGLSFIESNLASATE 153
 DB 87 HEDGSLVNDHHDH--QSLKDVLSSTLASILALKKMGIGERQINKGLQFIELNSALVTD 143
 QY 154 KSQPSPIGDIITPGLELVAK--NIDINLLSKOTDFSLMLHKEELKQKCHSNEM----D 207
 DB 144 ETIQKPTGFDIIPFGIKYARDNLNLTPIGSEVVDD--MIRKDLDLK-CDSEKFKSGRE 200
 QY 208 GYLAVISEGLNLYDMWVKYQKMGVSPNSPATAAAPHNQPCCLNLYNLSLDKFG 267
 DB 201 AVLAVLEGTRNLKWDNLIVKYQKNGSLFDSFATTAATQFGNDGCLRYLCSLLQKPE 260
 QY 268 NAVPTVPHDLFRLSMVDITELGISHHPRVEIKVLDITYRCWVERDEQIFMVDVTC 327
 DB 261 AAVPSVYPPDQVRLSIIVTLESIGDRDPKTEIKSLDITYRWLGRDGEICLDLATCA 320
 QY 328 LAPRLIRINGEYSPDLAETNELA-----LXDEYAALTYHASHLYQEDLSSGK 379
 DB 321 LAPRLLAHGYDVSVDLPKPFABESGSDTLEGVKNFTSVLELFAAQ-SYPHESALKK 379
 QY 380 QILKSADFLKEIITST--DSNRLSKLIHKEVENALKFPINTGLERINTRRNIQYN-VDNT 436
 DB 380 QCCWTQKYLEMEISSLWVKTSVRDKYLKKEVEDALAFPSYASLERSDHRKILINGSVENT 439
 QY 437 RILKTYHSSNISNTDYLRLAVEDFTTCQIYREELKGLRWVVENKLDQLKFAKQTAY 496

DB 440 RVTKTSYRLHNICTSDILKLAVDDEFQCSIHREMERLDRWIVENRIQLKFAKQKLA 499
 QY 497 CYFSAATILSPPELSDARISWAKNGILLTVVDDFDIGDTIDELTNLIQCVEKMNVDVK 556
 DB 500 CYFSAATILSPPELSDARISWAKNGILLTVVDDFDIGDTIDELTNLIQCVEKMNVDVK 559
 QY 557 DCCSEHVRILFLALKAICWIGDEAFKWQARDVTSHVIQTWLELMNMLREAINTRDAY 616
 DB 560 EYSSEHVEIIFSVLRDITILETGKAFYQGRNVTHIVKIVLWDLKSLMLREAEWSDKST 619
 QY 617 PTINMYMENAYVSPALGPVIVKPAIVFVGPKLSEBIESESEYHNLFKLNSTQGRLLNDIHS 676
 DB 620 PSLEDYMENAYISFALGPVILPATYILIGPPEKTVDSHQYNQLYKLVSTWGRLLNDIOG 679
 QY 677 FKREFKEGKINAVLHLSNGESGKVEEVEEEMMMKMKRKLMLKLIPEENGSIIVPRAC 736
 DB 680 FKRESAEGKINAVLHLSNGESGKVEEVEEEMMMKMKRKLMLKLIPEENGSIIVPRAC 739
 QY 737 KDAPWNVCHLVNFFYANDDGTGTNTILPTVKDIIYNPLVLVNNEN 780
 DB 740 KEAFLKMSKVLNLFYRKDDGFTSNDMLSLVKSIVIEPVSLQES 783
 RESULT 7
 AAW14055
 ID AAW14055 standard; protein; 789 AA.
 XX AC AAW14055;
 XX DT 17-OCT-2003 (revised)
 XX DT 27-JUN-1997 (first entry)
 XX DE Pumpkin ent-kaurene B synthase.
 XX Ent-kaurene B synthase; plant growth; seed growth; transgenic plant;
 XX pumpkin.
 XX Cucurbita maxima; L cv. Riesenmelone gelb vernetzt.
 XX EP768381-A2.
 XX 16-APR-1997.
 XX 07-OCT-1996; 96EP-00116005.
 XX 09-OCT-1995; 95JP-00261147.
 XX (RIKA) INST PHYSICAL & CHEM RES.
 XX Kamiya Y, Yamaguchi S;
 XX WPI; 1997-214817/20.
 XX N-PSDB; AAT60339.
 XX Pumpkin ent-kaurene B synthase polypeptide - used for the regulation of
 XX plant and seed growth.
 XX Claim 1; Page 13-19; 24pp; English.
 XX Pumpkin ent-kaurene B synthase (AAW14055) catalyses the conversion of
 XX copalyl pyrophosphate to ent-kaurene and has important functions in the
 XX growth of plants and seeds. The enzyme has been purified from pumpkin
 XX endosperm, and primers based on isolated peptides (see also (see also
 XX AAW14056-60) were utilised in the isolation of an ent-kaurene B synthase
 XX cDNA clone (AAT60339). The enzyme can be produced in transformed host
 XX cells, and DNA encoding the enzyme can be used to produce transgenic
 XX plants of increased height. (Updated on 17-OCT-2003 to standardise OS
 XX field)
 XX Sequence 789 AA;
 Query Match 48.9%; Score 2012.5; DB 2; Length 789;

| | | |
|----|--|--|
| XX | Sequence 740 AA; | |
| SQ | Query Match | 37.4%; Score 1537.5; DB 8; Length 740; |
| | Best Local Similarity | 41.7%; Pred. No. 2.2e-116; |
| | Matches 325; Conservative 142; Mismatches 243; Indels 69; Gaps 16; | |
| QY | 20 ALSAHTASTSGGQNTNPLIIDTKRIQKFNVEISVSSTAYTAWVPS-PNSPK 78 | |
| DB | 5 AMRHSSSSSEGGAAATAARSVREKL--QLAPSPSPSPYDTAWVAMPALRRGGG 62 | |
| QY | 79 SPFCPECLNWLNNQNDGSM--GLVNHHTNHP--LLKDSLSSTLACIVALKRNVGE 134 | |
| DB | 63 GRFPQCVAWIQENQDGSWRHAAAHQQLGSSPEIVTERDLSSTLACVALARWDAGS 122 | |
| QY | 135 DQNKGLSTIENLASATKSPSPGPDII--PGLLEYAKNLDINLIS-----KQTFP 187 | |
| DB | 123 EHVRRGLQFIGNMSVAMDQTAAPASGSGVSPFAAMLRMAMEGLEVPVAVSQADVRDRA 182 | |
| QY | 188 SLMLHKRELQKRECHSNEMDGYLAVISEGLNLYDNMVKYQMKNGSVFNSPSATAAF 247 | |
| DB | 183 GVI-----CHGRTE-YTAVYSEGLNIGNWNEVNFQKNGSLFNSPYTTAAL 231 | |
| QY | 248 INHONGCLNLYNSLDFKGNVPTVYPHDLFIRLSMVDTIERLGHSHFRVEIKNVLDE 307 | |
| DB | 232 VHYNDAKALQYLDMLDKGSAVPAAYPANIQSLQVWVDLEKMGISRHFVGEIKSILDM 291 | |
| QY | 308 TTRCWVERDQIFMDVVTCALAFRLIRINGEYVSPPLAEIT-----NELA--LKDEYA 359 | |
| DB | 292 TYSCKQRDEEIVLDNQTGMAFRMLRMNGYDSSDELSHFSEPPSFHNSLOGLYNDTRS 351 | |
| QY | 360 ALETHASHILYQEDLSGGQILKSADFLKEIISTDSNRLSKLIHKEVENALKFPTNTGL 419 | |
| DB | 352 LLELHKASKV-----STAEKEVEYALEFPFTTIL 380 | |
| QY | 420 ERINTERNIQLNVNDTRILKTYTHSNTSDTYLRLAIVEDFTYCOQSIYREELKGLRWV 479 | |
| DB | 381 DRLDHKRNIEHPDITSSQMLETAYLPCH-SNEEIMALGVDRDFSSQFIFQEEIQQLNSWV 439 | |
| QY | 480 VENKLDQFKARQKTAFCVSVAAATSSPBLSPARISWAKNGILITVDDPFDIGTIDE 539 | |
| DB | 440 KESRLDQLFQAROKLDYFYSAAATFTPELSDVRILNAKNGVLLTVDVDFDVGSKKE 499 | |
| QY | 540 LTWLIQCVKWNVDVKDCSEHVRILFLTALDKAICWIGDEAPKQARDVTSHVITWLE 599 | |
| DB | 500 LEMVALVEKNDKDKTEYSEVSEVFSAYITSTQLGSMASVQGRDVTXHLVIWQE 559 | |
| QY | 600 LMNSMLREAITWTRDAYPTLNEYMENAYVSFALGPVYKPAIYFVGPKLSEIYESSEYHN 659 | |
| DB | 560 LLRSMTVEVWRQSRVPTAEYEMENAVVTFALGPVLPALYLVGPKIPDSVIRSQECSE 619 | |
| QY | 660 LFKLMSTOGLINDIHSFKREFEKGKINAVLHLSNGESGKVB--EEVVEEMMMKIKRK 718 | |
| DB | 620 LFRIMSKGRLNDVQSYEREGSQGLNSVSL--LALHSGGSVSMEEAVKQIQPIEKRR 678 | |
| QY | 719 ELMKLIFEENGSTVPRACKDAFNMCHLVNFFYANDDGFNTILDTVKD--IYNPLVL 776 | |
| DB | 679 ELLKLVSRGGA-VPRCPRELFWSMCKVCHFFISGGDGFSPSTAKALDAVIEPLNL 736 | |

RESULT 9

ADM94224

ID ADM94224 standard; protein; 732 AA.

XX

AC ADM94224;

XX

DT 17-JUN-2004 (first entry)

XX

DE Corn Ent-Kaurene Synthase B protein #1.

XX

KW Corn; plant; plant metabolism;

KW GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate synthase;

KW Dioxxygenase; Ent-Kaurene Synthase A; GA-20 oxidase;

Gibberellin 3-beta hydroxylase; Riboflavin specific deaminase; Ethylene response factor; Acyl-CoA thioesterase II; ABC transporter GCN20-like; P-glycoprotein I; P-glycoprotein 3; P-glycoprotein ATPGP; P-glycoprotein HVMR2; ABC transporter; PMP70 ABC transporter; MRP4 ABC transporter; transgenic. Zea mays.

US6677502-B1.

13-JAN-2004.

12-JUL-2000; 2000US-00614912.

12-JUL-1999; 99US-0143401P.

12-JUL-1999; 99US-0143412P.

30-JUL-1999; 99US-014650P.

15-DEC-1999; 99US-0170906P.

21-DEC-1999; 99US-0172946P.

21-DEC-1999; 99US-0172959P.

(DUPO) DU PONT DE NEMOURS & CO E I.

(PION-) PIONEER HI-BRED INT INC.

Allen SM, Kinney AJ, Rafalski JA, Orozco EM, Miao G, Famodu OO; Lee J, Lohman KN, Rendina AR, Sakai H, Weng Z, Caimi PG, Fang Y; Shen JB, Zoughi IL, Anderson SL, Shi J, Lu G, Helentjaris TG; Li CP;

WPI; 2004-088430/09.

N-PSDB; ADM94223.

New isolated GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate synthase nucleic acid and proteins, useful for creating transgenic plants where polypeptides are present at higher or lower levels.

Example 3; SEQ ID NO 24; 186pp; English.

The invention relates to an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate synthase activity, appearing as ADM94266. Also included are a vector comprising the polynucleotide, a recombinant DNA construct comprising the polynucleotide operably linked to at least one regulatory sequence, transforming a cell with the polynucleotide, a cell comprising the recombinant DNA construct, producing a plant (comprising transforming a plant cell with the polynucleotide and regenerating a plant from the transformed plant cell), a plant comprising the recombinant DNA construct and a seed comprising the recombinant DNA construct. Also disclosed as new are nucleic acid (cDNA) fragments and assembled contigs encoding plant metabolism proteins chosen from GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate synthase, Dioxxygenase, Ent-Kaurene Synthase A, GA-20 oxidase, Gibberellin 3-beta hydroxylase, Riboflavin specific deaminase, Ethylene response factor, Acyl-CoA thioesterase II, ABC transporter, GCN20-like, P-glycoprotein I, P-glycoprotein 3, P-glycoprotein ATPGP, P-glycoprotein HVMR2, ABC transporter, PMP70 ABC transporter and MRP4 ABC transporter. The nucleic acid fragments may be used to create transgenic plants where the polypeptides are present at higher or lower levels than normal or in cell types or in developmental stages in which they are not normally found. The polynucleotides can be used as probes for genetically and physically mapping genes and as markers for traits linked to those genes. The nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP). It can also be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. The nucleic acid fragments may also be used as hybridisation probes against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer. The peptides can also be used to immunise animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. Antibodies are useful for detecting the polypeptides in situ in cells or in vitro in cell extracts. The present sequence represents a plant metabolism protein (or

QY 456 LAVEDFYTCOSYREBELKGLRWVENKLDQKFAKQKTAICYFVSVAATLSSPELSDARI 515
DB 552 LAKLDFNMVQALHOKETQHIVSWRESGFNDLTFTQRPVEMYFVSVMSPPEFAACRI 611
QY 516 SWAKNGILTVVDFFDIGTIDELTNLIQCVEKNVDVDKCCSEHVRILFLALKDAIC 575
DB 612 AVAKTSCLAVILDDLYDTHGSLDDKLFSFSEARVRWDISVLDSVRDNLQKVCFLGLYNTVN 671
QY 576 WIGDEAFKQWQARDVTSHVITQWLELNMNMLREAIWTRDAYVPTLNEYMENAYVSPALGPI 635
DB 672 GFQKGLKEQGRDVLGKRLKWEGLLASVTKAEWSAAKYVPTFNEYVENAKVSIATATV 731
QY 636 VKPAIFYVGPKEISEEIVSESEYHNLFF-KLMSTQGRLLNDIHSFKBEFKGL-NAVALHL 693
DB 732 VLNSIFTGELLFDYILQVDDLRKSPFLHVSITGRLLNDITKYQAEARNRGELVSSVQCYM 791
QY 694 -SNGESGKVEEVEEVEEMMMIKMKELMKLIFEENGSIVPACKDAFNMCHVILNFFYA 752
DB 792 RENPEC--TEEALSHVYGIIDNALKEL-NWELANPASNAPLCVRELLFTARVMQLFYM 848
QY 753 NDGPF--TGNITLDTVKDIYNPL 774
DB 849 YRDGFGISDKEMKDHVSRTLFDPV 872
RESULT 11
ADM98978
ID ADM98978 standard; protein; 873 AA.
AC ADM98978;
XX ADM98978;
DT 01-JUL-2004 (first entry)
XX
DE Diterpene synthase polypeptide #15.
XX
KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
KW diterpene synthase; defence toxin; volatile defensive signal;
KW pollinator attractant; photoprotectant; enzyme.
XX
OS Ginkgo biloba.
XX
XX US2004072323-A1.
XX
XX 15-APR-2004.
XX
XX 07-JAN-2002; 2002US-00041018.
XX
XX 05-JAN-2001; 2001US-0259880P.
XX
XX (MATS/) MATSUDA S P T.
XX (HART/) HART E A.
XX
XX Matsuda SPT, Hart EA;
XX WPI; 2004-373921/35.
XX
XX New unicellular organisms comprising exogenous nucleic acids encoding a
XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
XX producing diterpenes and diterpene precursors.
XX
XX Claim 31; SEQ ID NO 398; 38pp; English.
XX
XX The invention relates to a unicellular organism for producing a diterpene
XX or diterpene precursor comprising an exogenous nucleic acid sequence
XX encoding a geranylgeranyl pyrophosphate synthase under the control of a
XX promoter operable in the organism, and an exogenous nucleic acid sequence
XX encoding a diterpene synthase under the control of a promoter operable in
XX the organism. The invention also relates to methods of producing a
XX diterpene or diterpene precursor and a method of isolating a diterpene
XX synthase comprising growing several cells in the presence of a
XX polyaromatic resin to make a cell/resin mixture, where at least one of
XX the cells further comprises at least one isolated and purified nucleic

CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 873 AA;
SQ
Query Match 26.1%; Score 1075; DB 8; Length 873;
Best Local Similarity 32.3%; Pred. No. 2e-78;
Matches 260; Conservative 159; Mismatches 329; Indels 56; Gaps 16;
QY 10 PLLTNSRPAALSAHTASTSHGGQTNPNLLIDITKERIQKQKNV---EISVSSYDFA 66
DB 86 PAVWKDDFDLSLTSPNSHATSKSSVDETINKRIQTLVKELQCMFQSGMGGEYNPSAYDFA 145
QY 67 WYAMVPSNPKSPCFPECLNMLINNOLDNSGWLNVHTNHNHPLLLKDSLSTLACIVA 126
DB 146 WVARIPSDGGAPOPPQTLQWLNQLPDGSWG-----EBCIFLAYDRVLTACLIT 199
QY 127 LKRWNVGDDQINKGLSFIBSNLAS-ATEKSQSPSIGFDIIFPGLLEAYAKNLDINLSKOT 185
DB 200 LKIWNKGGDIQVQGVFEVFEKHEEMKDEADNRPSPGVFVFFAMLDKSLGLDLPYHLP 259
QY 186 DFSMLMKRELQKQKCHSNEMDGY---LAYISEGLNLYDMNWKYKQKNVSPSPSA 242
DB 260 FTSQIHQKQKLOKIPNLVHNHQTALLYSLGLEQDVVDVQWDEITNLQSRDGSFLSPAS 319
QY 243 TAAAFINHQNPCLNYINSLDKFGNAVPTVPHDLFIRLSMVDITIERIGISHHFVEIK 302
DB 320 TACVFNHTQNKRLHFLNVLKFGDYVPCHPDLDFERLWAVDTVERLGDYFKKEIK 379
QY 303 NVLDETYRCW-VERDE-----QIFMDVVTICALAFRLRLRINGEYVSPDPLAIBTNE----- 351
DB 380 ESLDYVYRYWDAERGVGWACNPIPDVDDTAMGLRILRLHGVNVSVDVLENFDEKGDFF 439
QY 352 -----LAKOEYAALETYHSHILYQEDLSGKQLKSA-----DELKEIISTD 395
DB 440 CFAGQTQIGVTDN---LNLRYCSQVCF-----PGEKIMBEAKTFTTNHQLQNALKNNAFD 491
QY 396 SNRLSKLIHKEVENALKFPINTGLERINTRNITQLYNDVNTIRILKTYTHSSNISNTDYLK 455
DB 492 KWAVKDLPGVEYAIKYPWHRSMPLREARSYIEQFSGNDVWLKIVYKMLYVSNKYLE 551
QY 456 LAVEDFYTCOSYREBELKGLRWVENKLDQKFAKQKTAICYFVSVAATLSSPELSDARI 515
DB 552 LAKLDFNMVQALHOKETQHIVSWRESGFNDLTFTQRPVEMYFVSVMSPPEFAACRI 611
QY 516 SWAKNGILTVVDFFDIGTIDELTNLIQCVEKNVDVDKCCSEHVRILFLALKDAIC 575
DB 612 AVAKTSCLAVILDDLYDTHGSLDDKLFSFSEARVRWDISVLDSVRDNLQKVCFLGLYNTVN 671
QY 576 WIGDEAFKQWQARDVTSHVITQWLELNMNMLREAIWTRDAYVPTLNEYMENAYVSPALGPI 635
DB 672 GFQKGLKEQGRDVLGKRLKWEGLLASVTKAEWSAAKYVPTFNEYVENAKVSIATATV 731
QY 636 VKPAIFYVGPKEISEEIVSESEYHNLFF-KLMSTQGRLLNDIHSFKBEFKGL-NAVALHL 693
DB 732 VLNSIFTGELLFDYILQVDDLRKSPFLHVSITGRLLNDITKYQAEARNRGELVSSVQCYM 791
QY 694 -SNGESGKVEEVEEVEEMMMIKMKELMKLIFEENGSIVPACKDAFNMCHVILNFFYA 752
DB 792 RENPEC--TEEALSHVYGIIDNALKEL-NWELANPASNAPLCVRELLFTARVMQLFYM 848

QY 753 NDGCF--TGNVILDTYKDIINPL 774
||| : : { :::
Db 849 YRDGFGISDKEMKHVSRTLDPV 872

RESULT 12
AAAY90858
ID AAAY90858 standard; protein, 868 AA.
XX AC
XX AAY90858;
XT XX
DT 25-AUG-2000 (first entry)
XX DE
DE Grand fir abietadiene synthase protein sequence SEQ ID NO:56.
XX SY
KW Synthase; protein co-ordinate data; active site; modification; terpenoid;
KW 3-dimensional coordinate; alpha carbon atom; plant; terpene synthase;
KW isoprenoid; breeding programme; fragrance; flavour; pheromone;
KW defensive agent; pigment; antitumour; steroid hormone;
KW signal transduction pathway; bile acid; affinity purification;
KW photoreceptor; enzymatic synthesis; nutrient supplement;
KW immunological reagent.
XX OS
XX Abies grandis.
XX OS
PN WO200017327-A2.
XX PD
PD 30-MAR-2000.
XX PF
PF 17-SEP-1999; 99WO-US021419.
XX PR
PR 18-SEP-1998; 98US-0100993P.
PR 22-APR-1999; 99US-0130628P.
PR 23-AUG-1999; 99US-0150262P.
XX PA
PA (KENT) UNIV KENTUCKY RES DEPT.
PA (SALK) SALK INST BIOLOGICAL STUDIES.
XX PI
PI Chappell J, Manna KR, Noel JP, Starks CM;
DR WPI; 2000-292839/25.
DR N-PSDB; AAA38937.
XX PT
PT Novel terpene synthase enzymes, useful for producing terpene
PT hydrocarbons, e.g. fragrances or antitumor agents, are derived from known
PT enzymes by specific amino acid alterations.
XX PS
PS Claim 112; Page 443-445; 450pp; English.

The present invention describes an isolated terpene synthase (I) comprising a region with at least 20% identity to region 265-535 of a 548 amino acid (aa) sequence (Ia), given in AAY90831. (I) contains nine alpha-carbon atoms (alphaC) that have interatomic distances, between each other, within tabulated ranges, have a centre point (within a sphere of radius 2.3 Angstrom) within tabulated ranges, and have an ordered arrangement of R groups (defining aa side chains), excluding specific tabulated arrangements (tables given in the specification). (I), and related enzymes, are used to produce a wide range of terpenoids (e.g. cyclic, acyclic, optionally hydroxylated), useful e.g. as fragrances, flavours, pheromones, defensive agents, pigments, antitumor agents, components of signal transduction pathways, precursors of steroid hormones and bile acids, as photoreceptors and as co-factor side chains. Some synthases with little or no catalytic activity (and nucleic acids encoding them) are used as controls in the analysis of products formed by enzymatic synthesis; as nutrient supplements; for affinity purification of isoprenoids; or to develop immunological reagents or nucleic acids for monitoring expression of terpene synthase or inheritance of the gene in plant breeding programs. The new synthases may produce novel terpene products. AAA38910 to AAA38938 and AAY90831 to AAY90859 represent sequences used in the exemplification of the present invention

Sequence 868 AA;

diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polycyclic aromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from the printed seqdata.uspto.gov/sequence.html.

Sequence 868 AA;

Query Match 25.8%; Score 1062.5; DB 8; Length 868;
Best Local Similarity 32.3%; Pred.No. 2.1e-77;
Matches 263; Conservative 148; Mismatches 339; Indels 61;

| Matches | 261; | Conservative | 148; | Mismatches | 339; | Indels | 61; | Gaps | 17; |
|---------|------|----------------------|--------------------------|-------------------------|----------------|--------|-----|------|-----|
| QY | 7 | IAGPLLTKSNRPAAL--- | SAIHTASTSHGGGNTPTNLII | DTTKBRIQKQFNV--- | EISV | 60 | | | |
| Db | 79 | LSSSTLVKRBFPPG | FWKDDLLSDTSSHKVAASDEKR- | IETLSEIKMFRCMG | VGETNP | 137 | | | |
| QY | 61 | SSYDTAWAMVSPNSPKSPCF | PCLNLWLNINQINDGSLWNHNNH | NHPLPKDSLSST | 120 | | | | |
| Db | 138 | SAYDTAWARIPAVDGSN | PHFPETVWILQNQLKDGSGW- | ----- | BGFYFLAYDRILAT | 191 | | | |
| QY | 121 | LACTIVALKRWVGDQ | INKGLSFESNLASATEKSQP- | SPIGEDIIIPGLLEYAKNLDIN | 179 | | | | |
| Db | 192 | LACIIITLWRTGETQ | VQKGIFFRTOAGKMEADSHR | PGFEIVFAPMLKEAKILGLD | 251 | | | | |
| QY | 180 | LLSKQTDFFSLMLKRE | LQKRCHSN---EMDGYLAIYSE | GLNLYDWNWVKYQKMGSV | 236 | | | | |
| Db | 252 | LPYDLPFLKQII | EKREAKLRIPDVLVYALPTT | LLYSLEGQEIWDQIMKLSQKGSF | 311 | | | | |
| QY | 237 | FNSPSATAAFINHQNP | CGCLNYLNSLDKFGNAVPTVY | PHDLFTRLGNWDTIE | LGISHH | 296 | | | |
| Db | 312 | LSSPASTAAVFMRTG | NKKCLDFLNFVLKFTGNHVP | CHYPYLPDLPERLWAVDT | VERLIGDRH | 371 | | | |
| QY | 297 | FRYEIKNVLDIETVRC | WVERD-----EQIFMDVVT | CALAFRLLRINGEYVSDP | DLAEITNE | 351 | | | |
| Db | 372 | FKEEIKALDYVYSHW | DERGIGWARENPVDIDT | TAMGLRILRLHYNVSDVL | KTRFDE | 431 | | | |
| QY | 352 | LALKDEYAA----- | LETYHSHILYQEDLSGKQILKSA- | ----- | DFLKE | 390 | | | |
| Db | 432 | ---NGEFFCFLGQT | QGVTDMLNVNRCSHVSF---- | PGETIMEAKLCTERYLRN | ALEN | 483 | | | |
| QY | 391 | IISTDSNRSLKLTHKE | VENALKPINTGLBRINTRRN | IQLYVNDTRILKTTYHSS | NIISN | 450 | | | |
| Db | 484 | VDAFDKWAFFKNIRGE | YALKYTPWHKSMRPLEARS | YIENYGPDDVWLKTYVM | PPYISN | 543 | | | |
| QY | 451 | TDYLRILAVEDFYCT | QSTLYREELKGLRWVVENK | LDLKFAROKTAYCYFVS | VAATLSPEL | 510 | | | |
| Db | 544 | EXYLELAKLDFNKVS | LHQLEQDLRWRWKSGETD | LNFTRRVTEITYFSPAS | FIFEPEF | 603 | | | |
| QY | 511 | SDARISWAKNGILTV | VDDPFDIGITIDELTNLI | QCQEKWNVDDKCCSEHVR | ILFAL | 570 | | | |
| Db | 604 | SKCREVTKTSNFTV | ILDDLYDAHGSLLDLKL | FTESVKRMDLSL-VDMQ | PQOMKICRVGF | 662 | | | |
| QY | 571 | KDAICWIGDEAFKWA | QARDVTSHVITQVLELMS | NMLREAIWTRDAYVPTL | NEYMENAYVSF | 630 | | | |
| Db | 663 | YNTFNDIAKEGRER | QGRDVLGYIQNWVKVQLE | AVTKEAWSEAKYVPSF | NEYENASVSI | 722 | | | |
| QY | 631 | ALGPYIKPALYVGP | KLSEIVESSSEYHNIF-K | LMSTQGRLLNDIHSFK | REFKEGKUNAV | 689 | | | |
| Db | 723 | ALGTVVLLISALF | TEGVLTDEVLSKI | DRESFLQMLGTLRLVND | TKTYQAEQGGV-AS | 781 | | | |

| | | | | |
|----|--|-----|--|-----|
| Qy | | 690 | ALHLNSGSGVK--EEVVEEMMMIMIKRKLMLKLIPEENGSIIVPRACKDAFNMCHVLN | 748 |
| | | | : : : : : : : : : : : : : : : : | |
| Db | | 782 | ALQCYNKHDPKISEEALQHYSVMNALEELNR---EFVNKIPDIYKRLVPETARIMQ | 838 |
| Qy | | 749 | FFVANDCGFT---GNLTIDTVKDIIYNPL | 774 |
| Db | | 839 | LFYMGGDGLTSLHSDMEIKEHVKNCLFPV | 867 |

RESULT 15
ABU09780

ID ABU09780 standard; protein; 868 AA.

AC ABU09780;

DT 06-AUG-2003 (first entry)

DE white fir abietadiene synthase.

Levopimaradiene synthase; ginkgolide; enzyme; *Escherichia coli*;
Candida albicans; *Kluyveromyces lactis*; *Saccharomyces*; white fir;
 abietadiene synthase.

XX Abies grandis.

AA PN US2002164736-A1

XX
PD
07-NOV-2002[illegible][illegible][illegible]

PA (SCHE/) SCHEPMANN H G.

PI Matsuda SPT, Schepmann HG;

DR WPI; 2003-352584/33.

PT New purified and isolated nucleic acid sequence encoding a
PT levopimaradiene synthase useful for producing pure ginkgolides in cells.
PT

PS Example 3; Fig 4; 37pp; English.

The invention relates to a purified and isolated nucleic acid sequence encoding a levopimaradiene synthase. The nucleic acid is used for the preparation of a ginkgolide in a cell, an expression vector, a unicellular organism (e.g. *Saccharomyces*, *Escherichia coli*, *Candida albicans* or *Kluyveromyces fragilis*), a yeast host cell, a plant (particularly *Ginkgo biloba*), a host cell, a transgenic plant or a seed of the transgenic plant, for the biosynthesis of ginkgolides and in molecular biology, molecular genetics, and organic chemistry. The nucleic acid is capable of producing large amounts of pure ginkgolides in a cost-effective manner, particularly in an organism capable of a high-yield ginkgolide-producing system. This sequence represents an abietadiene synthase polypeptide which shows similarity to the levopimaradiene synthase of the invention

Sequence 868 AA:

| | | | | |
|-----------------------|--------|--------------------|-------|-------------|
| Query Match | 25.6%; | Score 1051.5; | DB 6; | Length 868; |
| Best Local Similarity | 32.0%; | Pred. No. 1.7e-76; | | |

OV 7 IASPI,I,TKSNRPAAI,----SAHTASTSHGGOTNPNTI,ITDTPYPTOVAFKAT,
PICH CA

79 LSSSTIYKPEEPDCEWKNPI IDETTCCHUYIA KCEVZ
::: | | | | : :: |
| | | | | : |

dp

QW 61 SCYPTETZLTMVING DNYCXYKDCEDDCT MWI TANYOT NINDCYOT TANTYMYTNTNCT

[illegible]

Mon Oct 18 07:21:11 2004

[illegible]

Search completed: October 17, 2004, 08:45:53
Job time : 212.188 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:27:39 ; Search time 8680.1 Seconds
(without alignments)
11721.026 Million cell updates/sec

Title: US-10-041-018-361
Perfect score: 2792
Sequence: 1 cccactcatctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 457.2 | 16.4 | 725 | 5 | BQ992296 QGF25D22. |
| 2 | 393.6 | 14.1 | 510 | 1 | AU292191 AU292191. |
| 3 | 337 | 12.1 | 665 | 1 | AI485628 EST243949 |
| 4 | 316.8 | 11.3 | 632 | 2 | BE434899 EST405977 |
| 5 | 310.8 | 11.1 | 843 | 7 | CO079063 GR_Ea41K |
| 6 | 305.2 | 10.9 | 590 | 1 | AI485623 EST243944 |
| 7 | 239.4 | 8.6 | 753 | 2 | BE658119 SM700004B |
| 8 | 233 | 8.5 | 821 | 6 | CB894537 EST647329 |
| 9 | 219 | 7.8 | 500 | 5 | EQ104995 S90295.e |
| 10 | 207.2 | 7.4 | 367 | 6 | CA824186 R37G11 tw |
| 11 | 205.6 | 7.4 | 1077 | 5 | BX838565 BX838565 |
| 12 | 195.8 | 7.0 | 543 | 6 | CF037118 QCG40h09. |
| 13 | 190.4 | 6.8 | 619 | 6 | CF197159 Ir64v11.5 |
| 14 | 189.6 | 6.8 | 567 | 6 | CF036221 QCG29C10. |
| 15 | 185 | 6.6 | 515 | 6 | CF037618 QCG9506.y |
| 16 | 183.6 | 6.6 | 596 | 6 | CA452452 Rxo-2_D07 |
| 17 | 174.6 | 6.3 | 582 | 6 | CF036148 QCG28B11. |
| 18 | 174.6 | 6.3 | 582 | 6 | CF036916 QCG38G06. |
| 19 | 172.8 | 6.2 | 439 | 1 | AI940878 sb79b02.y |
| 20 | 172.2 | 6.2 | 558 | 1 | AU252376 AU252376 |
| 21 | 171.6 | 6.1 | 443 | 4 | BE49839 HVSNE1001 |
| 22 | 164.6 | 5.9 | 808 | 6 | CB621697 OS11Ea07M |
| 23 | 164.4 | 5.9 | 646 | 7 | CF418783 USDA-FP 1 |
| 24 | 164.4 | 5.9 | 685 | 2 | AW738064 EST339491 |

| | | | | | |
|----|-------|-----|-----|---|--------------------|
| 25 | 163.2 | 5.8 | 529 | 6 | CA914596 |
| 26 | 160.6 | 5.8 | 197 | 1 | AU294693 |
| 27 | 158 | 5.7 | 526 | 6 | CB899647 PCBP02741 |
| 28 | 157.4 | 5.6 | 788 | 6 | CB623234 OS11Ea10N |
| 29 | 155.4 | 5.6 | 575 | 4 | BM318434 P11_13_E1 |
| 30 | 154.8 | 5.5 | 547 | 2 | AW617347 EST323758 |
| 31 | 154 | 5.5 | 699 | 5 | BO766175 BBR008.SQ |
| 32 | 152.8 | 5.5 | 825 | 5 | CB623957 OS11Ea12B |
| 33 | 146.2 | 5.2 | 555 | 2 | AW617193 EST323604 |
| 34 | 145.6 | 5.2 | 616 | 2 | AW616201 EST307240 |
| 35 | 145.6 | 5.2 | 618 | 2 | AW616591 EST323002 |
| 36 | 145.6 | 5.2 | 620 | 2 | AW616079 EST296838 |
| 37 | 145.6 | 5.2 | 625 | 2 | BE353211 EST400343 |
| 38 | 145.6 | 5.2 | 642 | 2 | AW616634 EST323045 |
| 39 | 145.6 | 5.2 | 705 | 4 | BE288466 EST348055 |
| 40 | 143.8 | 5.2 | 614 | 2 | AW616902 EST323313 |
| 41 | 142.6 | 5.1 | 689 | 2 | BE585476 EST46PSP6 |
| 42 | 142.4 | 5.1 | 556 | 4 | BG126452 EST472098 |
| 43 | 141.4 | 5.1 | 452 | 4 | BM501430 PAC000000 |
| 44 | 140.2 | 5.0 | 718 | 4 | BG837019 Zm08_1060 |
| 45 | 139.2 | 5.0 | 531 | 6 | CA646242 wreln.pk0 |

ALIGNMENTS

RESULT 1
BQ992296
LOCUS BQ992296 725 bp mRNA linear EST 21-AUG-2002
DEFINITION QGF25D22.yg.ab1 QG EFGHU lettuce serriola Lactuca sativa cDNA clone
QGF25D22, mRNA sequence.
ACCESSION BQ992296
VERSION BQ992296.1 GI:22411831
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
1 (bases 1 to 725)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
Ellison,P., Kolkman,J., Shabugh,M.S., Livingson,K., Zhou,Y.,
Lai,Z., Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmunsen Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QGF25 row: D column: 22.

FEATURES

source
1. 725
Location/Qualifiers
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGF25D22"
/lab_host="E.coli"
/clone_lib="QG EFGHU lettuce serriola"
/note=Vector: pBRCNASFIAB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize

size bias. Details of each source of RNA and library construction can be obtained at <http://cgdb.ucdavis.edu/>
TAG TISSUE=germinating seeds
TAG_LIB=QG EFGHU lettuce serriola
TAG_SEQ=TCTGTGGGG

ORIGIN

Query Match 16.4%; Score 457.2; DB 5; Length 725;
Best Local Similarity 79.9%; Pred. No. 7.1e-95; Mismatches 138; Indels 1; Gaps 1;
Matches 551; Conservative 0;

QY 1362 AACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAAATGCTCTTAAAGTTCCCTATCAAT 1421
DB |||||
QY 1422 ACGGTTTAGACGCAATAAAGCTAGACGAATATACAGCTTTACATGTAGACATACA 1481
DB |||||
QY 1482 AGAATTTCTGAAACTACATATCACTCATCAATATATTAGTAACAATGATTAACCTTAAGTTG 1541
DB |||||
QY 1542 GCTGTTGAAGATTTCACACCTGCCAATCTATTATCGTGAAGAATTAAGAAGCTTTGAA 1601
DB |||||
QY 1602 AGGTGGGTGTAGAGAAATGATTTGGACCACTCAAGTTGCTAGGAAAGACCGCTAC 1661
DB |||||
QY 1662 TGTATTCTCTGCTGCAACACTTTCGTCTCCCGAATTAATCAGATGCGCTATTCTCA 1721
DB |||||
QY 1722 TGGGCCAAATAATGGCATAATTAACACAGTAGTTGATGACTTTTGTGATATCGGTGTGACA 1781
DB |||||
QY 1782 ATCGATGAATTGACCACTGATTCATCTGTTGAAAATGGAATGATGATGTCGACAAG 1841
DB |||||
QY 1842 GATTGTTGTCAGACATGTTCCGATTTTATTTTATAGCATTAAGAATGCAATCTGTGG 1901
DB |||||
QY 1902 ATTGGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAACATGCAATGTTATTCAAACT 1961
DB |||||
QY 1962 TGGTGGAACTAATGAATAGTAGTTGAGAGAGCTATATGGAAGAGATGCTTATGTG 2021
DB |||||
QY 2022 CCAACATTAATGAATATATGAAAACGCT 2051
DB |||||
QY 2051 TCCACATAAATGAATATATGAAAATGGT 725

RESULT 2

AU292191 510 bp mRNA linear EST 04-DEC-2002
LOCUS AU292191 zinnia cultured mesophyll cell equalized cDNA Zinnia
DEFINITION zinnia cultured mesophyll cell equalized cDNA Zinnia
ACCESSION AU292191
VERSION AU292191.1 GI:24252699
KEYWORDS EST.
SOURCE Zinnia elegans
ORGANISM Zinnia elegans
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Asteroideae;

REFERENCE
AUTHORS

Heliantheae; Zinnia.
1 (bases 1 to 510)
Demura,T., Tashiro,G., Horiguchi,G., Kishimoto,N., Kubo,M.,
Matsuoka,N., Minami,A., Nagata-Hiwatashi,M., Nakamura,K.,
Okamura,Y., Sassa,N., Suzuki,S., Yazaki,J., Kikuchi,S. and
Fukuda,H.

TITLE

Visualization by comprehensive microarray analysis of gene
expression programs during transdifferentiation of mesophyll cells
into xylem cells

JOURNAL

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)

Contact: Taku Demura

Morphogenesis Research Group

RIKEN Plant Science Center

1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9605

Fax: 81-45-503-9573

Email: demura@postman.riken.go.jp

This clone was obtained at our laboratory.

Seq primer: M13 forward.

Location/Qualifiers

FEATURES

source

1..510

/organism="Zinnia elegans"

/mol_type="mRNA"

/cultivar="Canary bird"

/db_xref="taxon:34245"

/clone="Z6921"

/tissue_type="mesophyll cell"

/clone_lib="zinnia cultured mesophyll cell equalized cDNA"

/note="Vector: pGEM-T easy; cultured in tracheary element
differentiation-inductive medium"

ORIGIN

Query Match 14.1%; Score 393.6; DB 1; Length 510;
Best Local Similarity 86.3%; Pred. No. 3.6e-80; Mismatches 69; Indels 0; Gaps 0;
Matches 435; Conservative 0;

QY 1430 AGAACGCATAAACAATAGACGAAATATACAGCTTTCATGTAGACATACAAAGATTCT 1489
DB 1 AGAACGCATAAATATCTCGACGGAACATCGAGCATTAATAATGTAGGTGAAACAAAGATTTT 60

QY 1490 GAAACTACATATCACTCATCAATATTAGTACACTGATTAACCTAGGTGGCTGTGTA 1549
DB 61 AAAAATACTATTTCGATCATCGAGTATTAGCAACAGGATTACCGAAGTTGGCAGTTGA 120

QY 1550 AGATTTCTACACCTGCCAATCTATTATCGTGAAGAATTAAGAAGCTTTGAAAGGTGGGT 1609
DB 121 AGATTTTACACCTGCCAATCTATTATCCGTGAAGAATTAAGAAGCTTTGAAAGGTGGGT 180

QY 1610 GGTAGAGATAAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTT 1669
DB 181 GATAGAGATAAATTTAGACGAGCTAAATTTGCTAGGCAAAAGACAGCCTACTGTTACTT 240

QY 1670 CTCTGTTGTCGCAACTTTTCTCCGAATTTATCAGATGCGCTATTTTCATGGGCCAA 1729
DB 241 CTCTGTTGTCGCAACTTTTCTCCGAATTTATCAGATGCGCTATTTTCATGGGCCAA 300

QY 1730 AAATGGCATATTAACTACAGTAGTTGATGACTTTTTTTTGTATATCGGTGGTCAATCGATGA 1789
DB 301 AAATGGCATCTTACTACAGTAGTTGATGACTTTTTTTTGTATATCGGTGGTCAATCGATGA 360

QY 1790 ATTGACCAACCTGATTCAATGTGTTGAAAATAAGGAATAGATGTGACAGAGATTGTTG 1849
DB 361 ATTGGCAATCTGATTCAGTGTGTTGAAAATAAGGAATAGATGTGACAGAGATTGTTG 420

QY 1850 TTCAGAGCATGTTCCGATTTTATTATTTAGCATTAAGAAGATGCAATCTGTTGGATTGGAGA 1909
DB 421 TTCAGAGCATGTTCCGATTTTATTATTTAGCATTAAGAAGATGCAATCTGTTGGATTGGAGA 480

QY 1910 TGAAGCTTTTAAATGGCAAGCGCG 1933

DB 481 GACAGCTTTTAAATGGCAAGCGAG 504

Db 361 GTTCGGAGAGTAGGCATTATATATTTTCGGCGCTTCACAGTACTATATAAGTAGGAGTTGGAG 420

Qy 1909 ATGAAGCTTTAAATGCGCAAGCGCGCATGTAACTAGCCATGTTATTCAAACCTTGGTTGG 1968

Db 421 ACAAGCACTAAGTGGCAAGCACGTAGTGTACAGACACATACTGATATTGGTTAA 480

Qy 1969 AACTAATGAATAGTATGTTGAGAGAAAGCTATATGGCAAGAGATGCTTATGTGCCAACAT 2028

Db 481 ATCTACTGAATGCTATGTTAAAGAGAAGCTGAATGGGCTAAAGATATGTCCGTGCCATCAT 540

Qy 2029 TAAATCAATATATGGAAGAGCTTACGTGCTCATTTTCATTTGCGATGGCCGATGTCAAGCCCG 2088

Db 541 TGGACAAATATATGGCCAAACGGTTATGTATCATTTGCCCTAGAGCAAAATTTTCCTTCGCG 600

Qy 2089 CTATTTTACTTTGTGGGGCCCAAAATATCAGAGAGATGTTGAAAGCTCTGAATATCA 2148

Db 601 CACTTTATTTGTTGACCCAAACTCCAGATGATGTTGTGCAACATCTCTGATTTCCACA 660

Qy 2149 ATCTA 2153

Db 661 GTTTA 665

RESULT 4

LOCUS BE434899

DEFINITION EST405977 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA clone cLEG24M17, mRNA sequence.

ACCESSION BE434899

VERSION BE434899

KEYWORDS 1 GI:94322742

SOURCE EST.

ORGANISM Lycopersicon esculentum (tomato)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 632)

REFERENCE Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D. Generation of ESTs from tomato fruit tissue, breaker stage Unpublished (2000)

AUTHORS Contact: CUGI

COMMENT Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

Location/Qualifiers

1..632

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cLEG24M17"

/tissue_type="Pericarp"

/dev_stage="breaker"

/lab_host="SOLR"

/clone_lib="tomato breaker fruit, TIGR"

/note="Vector: pBluescriptSkmCladap; Site 1: EcoRI; Site 2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

ORIGIN

Query Match 11.3%; Score 316.8; DB 2; Length 632;

Best Local Similarity 68.8%; Pred. No. 2.3e-62;

Matches 435; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 1405 TTAAGTTCCTATCAATACCGGTTTAGAACGCAATAACACTAGACGAATATACAGCTTT 1464

Db 1 TTAAGTTTCCTTCTCATGCGAAACATAAAACGGGTAGCCCAACAGGAGAAAATATAAAGCACT 60
QY 1465 ACAATGTAGACATACAGAATCTGAAAACCTACATATACATCATCAATATATAGTAACA 1524
Db 61 ATGATGTAGATAATAACAAGAGTCTGAAAACCTTCATATAGTTTCATCAAAATTTTGGCAACA 120
QY 1525 CTGATTACCTAAGTGTGGCTGTGTGAAGATTCTACACCTGCCAATCTATTTATTCGTGAAG 1584
Db 121 AAGATTCTTAACCTGGCGGTAGAAAGACTTCACCTTTGCCAATCTATCCACCGCAAG 180
QY 1585 AATTAAGGTCTGTGAAAGGTGGGTGAGAGATAAGTTGGACCAAGCTCAAGTTTGTCTA 1644
Db 181 AATTAAGCACTGTGAGAGGTGGTTAAACGCAAAACAGATTGGACAAGCTGAAGTTCTGTCA 240
QY 1645 GGCAAAAGACCGCTACTGTATTTCTCTGTGCTGCAACACTTTTCTGCTCCGGAATTA 1704
Db 241 GAGAGAGATCTGCATACCTGCTACTTTCTGCTGCGACGCCAATTTTCAACCTGGAATAT 300
QY 1705 CAGATCGCGGTATTTTCATGGGCCCCAAAATGGCAATTAACACAGTAGTTGATGACTTTT 1764
Db 301 CTGATCCCGCATGTCTATGGGCCAAGATGGTGTACTTACTACAGTGAATGATGATTTCT 360
QY 1765 TTGATATCGGTGTGTAACAATGATGAAATGACCAACCTGATTCATGTTGAAAAATGGA 1824
Db 361 TTGATGTAGAGGTTCTATGGAAGAATTGAACCAACCTAATTCGTGTTTAAAGAGTGCG 420
QY 1825 ATGTAGATGCAAGAGATTTGTTTCAGAGCATGTTCCGATTTATTTTATAGCAATTA 1884
Db 421 ATGTAGATGTAACATGATGTTCTGTCGAGAGAGTAGGCATATATTTTTCGCGCTTC 480
QY 1885 AAGATCAATCTCTGATGAGATGAAGCTTTTAAATGGCAAGCGCGCATGTAACTA 1944
Db 481 ACAGTACTAATGAGATGGAGACAAAGCATCTAAGTGGCAAGCACGTAGTGTACAA 540
QY 1945 GCGATGTTATCAAACTGGTTGGAACATAATGATAGTATGTTGAGAGAAGCTATATGA 2004
Db 541 GACACATAACTGATATTTGGTTAAATCTACTGAATGCTATGTTAAGAGAAGCTGAATGG 600
QY 2005 CAAGATGCTTATGTGCGCAACATTAATGAA 2036
Db 601 CTAAGATATGTCGGTGCCATCATTTGGACAAA 632

RESULT 5
LOCUS CO079063
DEFINITION GR_Ea41K03.r GR_Ea Gossypium raimondii cDNA clone GR_Ea41K03 3', mRNA sequence.
ACCESSION CO079063
VERSION CO079063.1 GI:48748544
KEYWORDS EST.
SOURCE Gossypium raimondii
ORGANISM Gossypium raimondii
REFERENCE 1 (bases 1 to 843)
AUTHORS Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C., Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and Wing,R.A.
TITLE Global assembly of Cotton ESTs
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
The University of Arizona
Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: <http://genome.arizona.edu>
Plate: 41 row: K column: 03.
Location/Qualifiers
1. .843
/organism="Gossypium raimondii"

/mol_type="mRNA"
/db_xref="taxon:29730"
/clone="GR_Ea41K03"
/tissue_type="whole seedlings"
/dev_stage="first true leaves"
/lab_host="DH10B"
/clone_lib="GR_Ea"
/note="Vector: pCMV.SPORT-6.1; Site 1: NotI; Site 2: EcoRV; Library made by Invitrogen with RNA supplied by Wendle lab. Directional cloned into NotI-EV. Colonies plated/picked by AGI. More glycerol clones held in -80."

ORIGIN

Query Match 11.1%; Score 310.8; DB 7; Length 843;
Best Local Similarity 63.8%; Pred. No. 6e-61;
Matches 537; Conservative 0; Mismatches 272; Indels 33; Gaps 3;
QY 939 CTTAATTTAAATTCACCTTTTGGACAAAGTTTGGTAATGTCAGTCCCAACAGTTTATCCT 998
Db 1 CTTCAATATCTCACTGCACCTTTAGACCCGTTTGAGAATGCAGTTCACACACTTCATCCT 60
QY 999 CATGATTTATTTATCCGACTTTCTATGGTTGACAAATTTGAAAGATTAGAAATTTCCAC 1058
Db 61 TTCCATGTTTTCGCCGACTTTGTATGTTGGAGACTGTTGAAAAGTTTGGAAATGGTCAG 120
QY 1059 CATTTCAGAGTGGAAATTTAAATGTTTATAGATGAAACATACAGATGTTGGGTGGAACGA 1118
Db 121 CATTTCAGGAGAAATTCACAGTGTATTAGATGAAACATATCGATGTTGGCTACAGGG 180
QY 1119 GATGACAAATATTCATGATGTTTAAACATGTCGTTTACGCTTTCGGTTATTTAAAGATC 1178
Db 181 GAGGAGGATATTTCTGGATCTGCCGACATGTCGATTTGGCTTTCAGAACACTACCTGTA 240
QY 1179 AATGGGTAGAAATTTCCCGACATCCATTTGGCTGAAATTAATAA----- 1222
Db 241 AAGCGTTATGATGTTTCCCTCAGAAGCCTTAGCTGGATTTGCTGGAAGAACACTTCTCAAT 300
QY 1223 -----TGAATTAGCTTTGAAAGACGAATATGACGCTCTTTGAAACATATCATCGGTGACA- 1276
Db 301 TCACITGGAGGATATTTGAGGATTTGGATGCTGCTGTTGAGCTATTTAGGGCTTCACAG 360
QY 1277 --TATATTATACAGAGAGATTTATCTTCTGGAAGAACAAATCTTGAAGTACGCTGATTC 1334
Db 361 ATGATCATACATCCAAATGAAACAACTCTCTGGAGAAACACATCTCATGGAACAGTCATTC 420
QY 1335 CTCGAAGATATATATCCACTGATTCGAACAGGCTTTCTAAATTAATTATTCACA----- 1386
Db 421 TTGAAACAGGAATTTATCCAAACACTTCTAAATGTGCATATAAAACATAAAATATATTATG 480
QY 1387 -AAGAGGTGGAATAATGCTCTTAAGTTCCCTATCAATACCGGTTTAGAAGCGCATAAACACT 1445
Db 481 CAGAAGTGATGATGCCCTTGAGTTTCCCATTTATGCAAGTTTGGAGCGCTAGTATAC 540
QY 1446 AGACGAAATATACAGCTTTTACAAATGATGAGCAATACAAAGATTTCTGAAAACATCATATC 1505
Db 541 AGGAGAAACATAGTGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 1506 TCATCAAAATATTAGTAACACTGATTAACCTTAAGGTGCGCTGTTGGAAGATTTCTACACTGC 1565
Db 601 TCTTTGAGTATTGGCAACAAAGATTTCTTAAGCTAGCAGTGGAGGACTTCAATGCGCTGT 660
QY 1566 CAATCTATTATTCGTGAAGAAATTTAAAGGTCTTGAAGGTGGGTGGGTAGAGATAAGTTG 1625
Db 661 CAATCTATATGCTGAAGAACTCAAAACACTCGAGAGGTGGGTTCAGAAAAGCGATG 720
QY 1626 GACCACTCAGTTTCTAGGCAAGACCGCTACTGTTATTTCTGTTGCTGCAACA 1685
Db 721 GACAGCTGAAGTTTTCAGGCGAAGAACTGGCATACTGTTACTTTTCTGAGCTGCAACA 780
QY 1686 CTTTGTCTCCGAAATTTATCAGATGCGGCTATTTTCATGGGCCAAAATGCGCATTAACCT 1745
Db 781 CTATGCCCTCTGAATTAICTGATGCTCGCTTAACATGGCTAAAATGCGGTGCTTACT 840

```

QY      1746 AC 1747
Db      841 AC 842

RESULT 6
AI485623
LOCUS   EST243944 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION
ACCESSION AI485623
VERSION   EST.
KEYWORDS  Lycopersicon esculentum (tomato)
SOURCE   Lycopersicon esculentum
ORGANISM Lycopersicon esculentum (tomato)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 590)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.W., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1..590
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED6120"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/clone_lib="tomato ovary, TAMU"
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

FEATURES
source
Query Match 10.9%; Score 305.2; DB 1; Length 590;
Best Local Similarity 69.8%; Pred. No. 1.1e-59;
Matches 412; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY      1489 TGAAGCTTCTACACCTGCGCAACTATTATTCGTGAGAGATTAAAGGTTCTGAAAGTGGG 1608
Db      1 TGAAGCTTCTACACCTGCGCAACTATTATTCGTGAGAGATTAAAGGTTCTGAAAGTGGG 1608
QY      1549 AAGATTCTTACACCTGCGCAACTATTATTCGTGAGAGATTAAAGGTTCTGAAAGTGGG 1608
Db      61 AAGATTCTTACACCTTGGCAACTATTCACCGCAAGAAATTAACAACTTGAGAGGTGGT 120
QY      1609 TGGTAGAATAAGTTGGACAGCTCAAGTTTGTCTAGGCAAAAGACCGCTACTGTTATT 1668
Db      121 TAACGCAAAACAGATTGGACAGCTGAAGTTCTGTCAGAGAGATCTGCATCTACT 180
QY      1669 TCTCTGTTGCTGCAACACTTTGCTCCCGAATTATCAGATGCGGTATTTTCATGGGCA 1728
Db      181 TTTCTGCTGCGCAACAAATTTTCAACCTGAACTATCTGATGCGCGCATGTTCATGGGCA 240
QY      1729 AAAATGGCATATTACTACAGTAGTGTGATGACTTTTTCATATCGGTGATCAATCGATG 1788
Db      241 AGAATGGTGTACTTACTACAGTAGTGTGATGATTTCTTTGATGTAGGAGGTTCTATGGAAG 300
QY      1789 AATTGACCAACTGATTCAATGTGTGTGAAAAAATGGAATGTAGATGTGCAAGGATTGTT 1848

```

```

Db      301 AATTGAACACCTAATTCGTGTTTAAAGAAGTGGAGTAGATGTGACCACTGATGCT 360
QY      1849 GTTCAGAGCATGTTCCGATTTTATTTTATTTAGCAATTAAGATCAATCTCTTGGATTGGAG 1908
Db      361 GTTCGAGAGAGTAGGCTATTATTTTCGGCGCTTCACAGTACTATAAGTGAGATTGGAG 420
QY      1909 ATGAAGCTTTTAAATGGCAAGCGCCGATGTAACTAGCCATGTTTAACTTGGTTGG 1968
Db      421 ACAAAGCATCTAAGTGGCAAGCACGCTAGTGTGTACAGACACACATACTGATATTTGGTTAA 480
QY      1969 AACTAATCAATAGTATGTTTGAAGAAGCTATATGACAAAGAGATGCTTATGTGCCAAT 2028
Db      481 ATCTACTGAATGCTATGTTTAAAGAAGCTGATGAGGCTTAAAGATATGTCGGTGCATCAT 540
QY      2029 TAAATGAATATATGAAACCGCTTACGTGTCTATTTGCAATTAGCCCGATT 2078
Db      541 TGGACAAATATATGCGCAACGGTTATGTATCATTTTGCCTAGGACCAATT 590

RESULT 7
BE58119/c
LOCUS   BE58119
DEFINITION 753 bp mRNA linear EST 24-MAY-2001
ACCESSION BE58119
VERSION   BE58119.1 GI:9984011
KEYWORDS  Glycine max (soybean)
SOURCE   Glycine max
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 753)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Corvelli,V.,
Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: AI940878 corresponding to Gm-cl010-1156 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
Location/Qualifiers
1..753
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-1532"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by

```

the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html>. Reracking was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Query Match 8.6%; Score 239.4; DB 2; Length 753;
Best Local Similarity 60.5%; Pred. No. 2e-44;
Matches 369; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 1892 AATCTGTTGGATGGAGATGAAGCTTTTAAATGCGAAGCGCGGATGTAAGTAACTAGCCATGT 1951
DB 748 AGTTNNTGANNNNAGANNNTCAGTCANNAGCAANGACGCAATGTGAAAACAATGT 689

QY 1952 TAFTCAAACCTTGCTTGGAACTAATGAATAGTATGTTTCAGAGAAAGCTATATGCAACAAGAGA 2011
DB 688 TATCAAAATTTGNNNGNAITTTGCTCCAGTCAATGTTTAGAGAAAGCTGAGTGGTTGAGAAC 629

QY 2012 TGGTTATGTCGAACAAATTAATGAATATATGGAAGAAAGCGTTAGTGTCAATTCATTTAGG 2071
DB 628 CAAGACTGTGCCAACAAATTTGCTGACTATATGAAAAATGCATATATCATTTTNCNTAGG 569

QY 2072 CCGGATTTCAAGCGCGCTATTTACTTTGTTGGGGCCCAATTAATCATGAGGAGATTTGTTGA 2131
DB 568 ACCAATTTGCTTCCGGCCNCTATNNGTTGGANNNAAGCTTTTCAGATGAATGCACAGA 509

QY 2132 AAGCTCTGAATATCATAATCTATTAAAGCTAATGAGCAGCAGCGGTGCGACTTTCTAAACGA 2191
DB 508 AATCATGAGTTGAACTATCTCTATAAGCTCATGACACATGTGGCGCTTCTCTTAATGA 449

QY 2192 TATCCATAGTTCAGAGGGAATTTAAGGAAGCAATTAAGCGCGGTAGCATTTGATTT 2251
DB 448 CATTCACAGTTTAAAGAGAGAAATCCGAGGAAGGAAATTTGAATGTGTGGCTTTGGGTAT 389

QY 2252 GAGTAACGAGAAAGTGGGAAGTGGGAAGAGAGTGTGTGGAGAGATGATGATGAT 2311
DB 388 TGCTCATGGAATGGAGTTATTTCTGCAAGAGATGCCACGAAGAAATGAAGGTTATGC 329

QY 2312 TAAAAACAAGAGAAAGAAATTAATGAATTAATTTTGAAGAAATTTGTTAGCATTTGTTCC 2371
DB 328 TGAAGAAAGAGAGAGAACTCTCTGAGATTAATTTTGCAGAAAGAAAGTGTAGTTCC 269

QY 2372 TAGAGCTTGAAGATGCATTTTGAACATGTCTCAGGTGTTGAATTTTTTTTACGCAA 2431
DB 268 CAGAGAGTGCAGAGATTTGTTTGGAAATGATAAAAGTGTTCACCTGTTTACATGAA 209

QY 2432 CGATGCGGTTTACTTGAACAACAGATTTCTTGATCTGTCAAGGACATCATTTTACAAACC 2491
DB 208 GGAATGGAATTTACTTCAACAGATGCACTCTAGTGTAAATGCAGTACTTTAAAAACC 149

QY 2492 GTTGGTGCTT 2501
DB 148 TGTCACTCTT 139

RESULT 8

CB894537
LOCUS
DEFINITION
EST647329 HOGA Medicago truncatula cDNA clone HOGA-31D19, mRNA
821 bp mRNA linear EST 24-APR-2003
sequence.
ACCESSION
CB894537
VERSION
CB894537.1 GI:30101706
KEYWORDS
EST.

SOURCE
ORGANISM

Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE

1 (bases 1 to 821)

AUTHORS

Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,

TITLE

Uterback,T., Cho,J. and Fraser,C.M.

JOURNAL

ESTs from roots of Medicago truncatula treated with

COMMENT

oligogalacturonides of DP 6-20

Unpublished (2001)

Contact: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMDU22TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtg gat CC).

FEATURES

source

1..821

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="HOGA-31D19"

/tissue_type="3 day old seedling roots"

/dev_stage="24 hours after treatment in the dark at 26 C

with 0.5 mg/ml oligogalacturonides (DP 6-20) in the

presence of 100 ug/ml Gentamicin"

/lab_host="XLOLR"

/clone_lib="HOGA"

/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Unizap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-Zap phage using Ex-assist

helper phage and propagated in SOLR cells."

ORIGIN

Query Match 8.3%; Score 233; DB 6; Length 821;
Best Local Similarity 62.9%; Pred. No. 6.3e-43;
Matches 403; Conservative 0; Mismatches 220; Indels 18; Gaps 2;

QY 1381 TTCACAAAGAGGTGGAAAATGCTTTAAGTTCCTATCAATACCGGTTTGAACGCAATAA 1440
DB 180 TTGATAATGAGTCAATGATGTTCTTAAAGTTTCCACATCATGCGAATTTAGAGCGTTGT 239

QY 1441 ACCTAGAGCAATATACAGCTTTTACAATGTACACATACAAGATTTCTGAAACTACAT 1500
DB 240 TAAACAGGAGATCAGTGGAGCATTAATGCGATGAACAGAGATTTTGAACATCCT 299

QY 1501 ATCACTCATCAATATTAGTAAACATGATTACCTAAGGTTGGCTGTTGAAGATTTCTACA 1560
DB 300 ACAGATCATGCATCTTGCAACCAAGAAATTTCTGAGCTAGCAGTTGAAGACTTCAACC 359

QY 1561 CTGCGCAATCTATTATTCGTGAAGAATTAAGAGTCTTGAAGGTGGTGGTGGTAGAGATA 1620
DB 360 TCTGCCAATCAATACAAAATAAAGAGCTGAAACAACCTTTCAGGTGGATTTGTGAAAGCA 419

QY 1621 AGTTGACAGCTCAAGTTTGTAGCAAGACCGCTACTCTTATTTCTCTGTTGCTG 1680
DB 420 AACTGACAACTAGACTTTTCCGAGCAGAAACTGGCATCTGTTACTTCTCAGTGCAG 479

QY 1681 CAACACTTTGCTCTCCGAAATTAATCAGATGCGGTATTTTCATGCGGCAAAATGGCATAT 1740
DB 480 CTACTCTTTATTTCCCGCACTTTCTGATGCTGCGATATCTTGGCGAAAAATGGGTAC 539

QY 1741 TAACACAGTAGTTGATGACTTTTTTTCATATCGTGTGTACATCGATGAATTCACCAACC 1800

```

Db      540 TCACACACTGTCGATGATTTCTTGATGTCGGGAGTTCTGAAGAAGAGCAAGTGAACC 599
QY      1801 TGATTCATGCTGTGA AAA-----TGAATGTAGATGTGACAGGA 1843
Db      600 TTATTCATGATGAGAAATATATTCCTAGTATTTAGTCGGATGTAGATGTCAATACTGT 659
QY      1844 TTGTTGTTCCAGAGCATGTTTCGGATTTTATTTTATTTAGCATTTAAAAGATGCAATCTGTGGAT 1903
Db      660 TTCTGTTCTGAGGCGTAAAGATATATTTCTGAGTTCGTAGCACAATTTGGAGAT 719
QY      1904 TCGAGATGAAGCTTTAAATGCGAAGCGCGCATGTAACTAGCCATGTTAT-TCAAAATT 1962
Db      720 TCGAGAGAAATCTGTTAGCGGCAAGCGGCAATGTGAAGACAATGTTATCAAAAATTT 779
QY      1963 GGTTCGAACTAATGATGATGTTGAGAGAGCTATATGG 2003
Db      780 GGTTAGATTGAGCGCTCTATGTTTACAGAAGCTGAGTGG 820

RESULT 9
LOCUS   BQ104995
DEFINITION 500 bp mRNA linear EST 16-APR-2002
          590295.e Rose Petals (Golden Gate) Lambda Zap Express Library Rosa
          hybrid cultivar cDNA clone gg0295.e 5', mRNA sequence.
ACCESSION BQ104995
VERSION   BQ104995.1 GI:20154657
KEYWORDS EST.
SOURCE    Rosa hybrid cultivar
          Rosa hybrid cultivar
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids I; Rosales; Rosaceae; Rosoideae; Rosa.
REFERENCE 1 (bases 1 to 500)
          Gureman,I., Shalit,M., Menda,N., Piestun,D., Dafny-Yelin,M.,
          Shalev,G., Bar,E., Davydov,O., Ovadis,M., Emanuel,M., Wang,J.,
          Adam,Z., Fichersky,E., Lewinsohn,E., Zamir,D., Vainstein,A. and
          Weiss,D.
          Rose Scent: Genomics Approach to Discovering Novel Floral
          Fragrance-Related Genes
          Plant Cell 14 (10), 2325-2338 (2002)
JOURNAL  Plant Cell 14 (10), 2325-2338 (2002)
COMMENT  Contact: Naama Menda
          Petal Genomics
          Faculty of Agricultural, Food and Environmental Quality Sciences,
          The Hebrew University of Jerusalem
          P.O. Box 12, Rehovot, 76100, Israel
          Tel: 972 8 9489 683
          Fax: 972 8 9468 265
          Email: shanama@agri.huji.ac.il
          Seq primer: T3 forward.
          Location/Qualifiers
            1..500
              /organism="Rosa hybrid cultivar"
              /mol_type="mRNA"
              /strain="Golden Gate"
              /db_xref="taxon:128735"
              /clone="gg0295.e"
              /tissue_type="Petals"
              /dev_stage="Young open flower at stage four"
              /clone_lib="Rose Petals (Golden Gate) Lambda Zap Express
              Library"
              /note="Vector: pBKCMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES
         source
         1..500
           /organism="Rosa hybrid cultivar"
           /mol_type="mRNA"
           /cultivar="Beaupre"
           /db_xref="taxon:3695"
           /dev_stage="two-month-old"
           /clone_lib="two-month-old roots from clone 'Beaupre'"
           /note="Organ: root; Vector: pRiplex2; cDNA library of
           two-month-old roots from Populus trichocarpa Torr.& Gray
           x deltoides Bartr. Ex Marshall (clone 'Beaupre'). The cDNA
           library was constructed from 1 ug of total RNA using the
           SMART cDNA synthesis kit (Clontech, Palo Alto, CA, USA)

ORIGIN
Query Match 7.8%; Score 219; DB 5; Length 500;
Best Local Similarity 65.8%; Pred.No.1e-35;
Matches 318; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY      1922 ATGGCAAGCGCGCATGTAAGTATGAGCATGTTATTCAAACTTGGTTGGAACTAATGAATAG 1981
Db      10 AAGCAAGCGCGCATGTGACAAATCAGTCATTCAGATTTGGTTGGATTTGCTCAAGGC 69
QY      1982 TATGTTGAGAGAGCTATATGGACAAGAGATGCTTTATGTGCGCAACATTAAATGAATATAT 2041

```

```

Db      70 TATGCTGAGGAGCTAGTGTGTTGATAAACAGTCAGCGCCACAACTGGAGGAATACAT 129
QY      2042 GGAACACGCTTACGTGTCATTTTCATTAGCCCGGATTTGTCAAGCCGGCTATTACTTTGT 2101
Db      130 GGAACATGCGTACGTCTCATTTTGCCTTGGGACCTAATGTCTCTTCAATCTCTATTGTT 189
QY      2102 GGGGCCCAATATCAGAGGAGATTTGTTGAAGCTCTGAATATCATATCTATTAAAGCT 2161
Db      190 AGGGCTTAAGCTCTCAGAGGATGCTGTACGGAGTTCAAGATACTCCATCTCTATAGACT 249
QY      2162 AATGAGCAGCGAGGTCGACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGA 2221
Db      250 AATGAGCACTTTTGGCGTCTTCTCAATGATGCAAGCTATAGAGGGAAGCTGCCGA 309
QY      2222 AGGCAAAATTAACGCGGTAGCATTTGCATTTAGTAAACGAGAAAGCTGGGAAAGTGAAGA 2281
Db      310 AGGGAAGCTAAATGCTGTAACATTTGCCATGATTCATGGCAATGTTAGTGTAGCGAGA 369
QY      2282 AGAGGTTGTGGAGGAGATGATGATGATGATTTAAACACAGAGAGAAAGATTAATGAATT 2341
Db      370 AGAGCCCATCAATGAGATGAAGAGCATTTATCCACAGTAAGAGGAGAGAACTGCAAGACT 429
QY      2342 AATTTTTCAGAAAAATGTTAGCATTTGCTTAGAGCTTGTAAAGATGCAATTTTGAACAT 2401
Db      430 AGTTTTCAGGACACAGGATAGTGTAGTTCCAGAGCTTGCAGAGATCTGTTTGAACAT 489
QY      2402 GTG 2404
Db      490 GGG 492

RESULT 10
LOCUS   CA824186
DEFINITION 367 bp mRNA linear EST 28-FEB-2003
          R37G11 two-month-old roots from clone 'Beaupre' Populus balsamifera
          subsp. trichocarpa x Populus deltoides cDNA 5', mRNA sequence.
ACCESSION CA824186
VERSION   CA824186.1 GI:28607735
KEYWORDS EST.
SOURCE    Populus balsamifera subsp. trichocarpa x Populus deltoides
          Populus balsamifera subsp. trichocarpa x Populus deltoides
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.
          1 (bases 1 to 367)
          Kohler,A., Delaruelle,C., Martin,D., Encelet,N. and Martin,F.
          The poplar root transcriptome: analysis of 7000 expressed sequence
          tags
          FEBS Lett. 542 (1-3), 37-41 (2003)
JOURNAL  Contact: Martin FM
COMMENT  Equipe de Microbiologie Forestiere
          Institut National de la Recherche Agronomique
          Centre INRA de Nancy, 54280 Champenoux, France
          Tel: +33 383 39 40 80
          Fax: +33 383 39 40 69
          Email: fmartin@nancy.inra.fr
          Insert Length: 367 Std Error: 0.00
          Seq primer: Forwat 5' AAGCGCGCATTTGTTGGTACCC.
          Location/Qualifiers
            1..367
              /organism="Populus balsamifera subsp. trichocarpa x
              Populus deltoides"
              /mol_type="mRNA"
              /cultivar="Beaupre"
              /db_xref="taxon:3695"
              /dev_stage="two-month-old"
              /clone_lib="two-month-old roots from clone 'Beaupre'"
              /note="Organ: root; Vector: pRiplex2; cDNA library of
              two-month-old roots from Populus trichocarpa Torr.& Gray
              x deltoides Bartr. Ex Marshall (clone 'Beaupre'). The cDNA
              library was constructed from 1 ug of total RNA using the
              SMART cDNA synthesis kit (Clontech, Palo Alto, CA, USA)
FEATURES
         source
         1..367
           /organism="Populus balsamifera subsp. trichocarpa x
           Populus deltoides"
           /mol_type="mRNA"
           /cultivar="Beaupre"
           /db_xref="taxon:3695"
           /dev_stage="two-month-old"
           /clone_lib="two-month-old roots from clone 'Beaupre'"
           /note="Organ: root; Vector: pRiplex2; cDNA library of
           two-month-old roots from Populus trichocarpa Torr.& Gray
           x deltoides Bartr. Ex Marshall (clone 'Beaupre'). The cDNA
           library was constructed from 1 ug of total RNA using the
           SMART cDNA synthesis kit (Clontech, Palo Alto, CA, USA)

```

according to the manufacturer's instructions. The resulting cDNA was packed into lambda phages using the Gigapack III Gold packaging kit (Stratagene, La Jolla, CA). The pTriplex2 phagemid clones in Escherichia coli were obtained by using the mass in vivo excision protocol according to the manufacturer's instructions (Clontech)."

ORIGIN

Query Match 7.4%; Score 207.2; DB 6; Length 367;
Best Local Similarity 73.9%; Pred. No. 5.2e-37;
Matches 263; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 1473 GACATACAGAAATCTGAAACATACATATCACTCAATCAATATTAAGTAACTGATTAAC 1532
DB |||||
DB 10 GATGATACAGAGATTTGAAACCTTCGTATCGTTGTTCCACTATTTGGTAACCGGATTTT 69
QY 1533 CTAAGGTTCGGCTGTTGAAGATTTCTACACCTGCCATCTATTATTCGTGAAGAAATTAATA 1592
DB |||||
DB 70 CTCAAATTCGCTGTAGAGACTTCAATATCTGCCAATCAATACAGCGTGAAGAAATTAATA 129
QY 1593 GGTCTTGAAGGTGGGTGTAGAGATAAGTTGGACAGCTCAAGTTTGGTGGCAAGAG 1652
DB |||||
DB 130 CATATCGAGGTGGGTGTTAGAGAGAGAGCTGGACAAGCTAAAGTTTGTCTAGGCGAGAAG 189
QY 1653 ACGGCTACTGTTATTTCTGTTGCTGCAACACTTTTGGTCTCCGAAATTTATCAGATGCG 1712
DB |||||
DB 190 GAGGCTACTGTTACTTCTGCTGCAGCAACTCTCTTGCCTCAGAACTATCTGAATGA 249
QY 1713 CGTATTTTCATGGCCCAAAATGCGATATTAACTACAGTAGTGTGATGACTTTTTTGTATATC 1772
DB |||||
DB 250 CGCATGTCATGCGCAAAATGTTGCTTACTACTGTTGTTGATGACTTCTTTGATGTT 309
QY 1773 GGTGTTACATCATGATGATGACCAACTGATTCATGTTGTTCAAAATGGAATGT 1828
DB |||||
DB 310 GGTGTTCTTGAAGAAGAACTGGTAAACCTTTATTGAATTTGATGAGAGGTGGGATGT 365

RESULT 11
BX838565
LOCUS
DEFINITION
Arabidopsis thaliana Flowers and buds Col-0 Arabidopsis
thaliana cDNA clone GSLTFB67ZB05 5PRIM, mRNA sequence.
BX838565
BX838565.1 GI:42532648
EST.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 1077)
Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C.,
Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,
Temple, G., Caboche, M., Weissenbach, J. and Salanoubat, M.
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
Unpublished (2004)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
The sequences are based on single pass reads.
Life technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences).
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/EST

http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

FEATURES
Location/Qualifiers
1..1077
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="GSLTFB67ZB05"
/tissue_type="Flowers and buds"
/clone_lib="Arabidopsis thaliana Flowers and buds Col-0"

ORIGIN

Query Match 7.4%; Score 205.6; DB 5; Length 1077;
Best Local Similarity 54.7%; Pred. No. 1.5e-36;
Matches 527; Conservative 0; Mismatches 414; Indels 23; Gaps 5;

QY 280 ATCCACATCAATCTGATCATTTGTTCCCTGAGTGTCTCAATTTGTTAAATAATCACTTAATG 339
DB 41 AGCTAACATGTAGAGCTTTTGAGCAACAAGAGAGAGATTAGGAAGATGTTGGAGAAAG 100
QY 340 TAGAAATTTCTGTTCTTCATATGACACAGCATGGGTAGCCATGGTCCCTTCTCCAAACT 399
DB 101 TGGAGCTTTCTGTTTCGGCTAGGATCTAGTTGGTGGTACATGGTTCATCACCAGCT 160
QY 400 CACCCAAATCGCCTTGTGTTCCCTGAGTGTCTCAATTTGTTAAATAATCACTTAATG 459
DB 161 CCCAAATGCTCCACTTTTCCACAGTGTGTAATGTTATTGGATAATGAAATCAAG 220
QY 460 ATGGTTTCATGGGTCTGTTGTTATTCACACTCAATATCAATCACTCCCGTTCCTTAAAGATT 519
DB 221 ATGGATGTTGGGAAATTGATA-----ACTATGAGGATGAATGTGTAAAGAGATG 271
QY 520 CTCATCTTCAACATTAGCATGTATTGTTGCATTAATAAAGATGGAATGTTGGGGAAGATC 579
DB 272 TGTATGATGTAGAGTGGGTAGTATGATAGGGTTAAGAGTGGGAAATGTTGGAAAGAG 331
QY 580 AATAAATAAAGTCTAAAGTTTAAATGAGTCAAACTCTGCTTCCAGTCTACATAAAGAGT 639
DB 332 AATAAAGAAGGTATGAGGTTTATTGAGTGAATTTGCGATTAGTGAATGATTTTTTTA 391
QY 640 AACCATCTCCATTTGTTGATCATATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 699
DB 392 TAGAGAAAGAGAGAGGTTTGTATTTATTTTGGTGGGATGATTAATTTGGTAGAGATT 451
QY 700 TGGACATAAAACCTCTTTCAAAACAAACAGATTTTAGTTTGTCTACATAAAGAGGAAAT 759
DB 452 TGAATATGAGATTAATTTGGGCTGAGAGTGGTGGATGAATGATAAGAAAAGAGAT 511
QY 760 TGGAGCAAAAAGATGCCATTTCAATGAGATG-----GATGATACCTTGGCGTAT 809
DB 512 GTGGATATTAAATGTTGATAGTGAAGATTTTGAAGGGGAAGAGAGATATGTGGGATAT 571
QY 810 ATCTCTGAAGGACTCGGTAAATTTATATGTTGGAATATGTTGGAAGAAATATCAGATGAAA 869
DB 572 GTTTTGAAGGGGAAAAGAAAAATAAAGATTGGGATTTGATTAATTAATTAAGAAAGAAA 631
QY 870 AATGGTTCTCTTTTCAACTCACCATCAGCAACAGCTGCTGCTTTTCAATTAATCATCAAAAT 929
DB 632 AATGGGTAATGTTTGAATATGAGGTTAAAGAGAGATGTTTAAATAGTTTGGGAAAT 691
QY 930 CTTGGTTGCTTAAATTAATTTAAATTTCACTTTTGGACAGAGTTTGGTAATGAGTCCCAACA 989
DB 692 GAAGGTTGTATAAGTTATATATGTTATGTTTGTGTTTGAAGAAATAGAGGGTGAAGTTGATTTA 751
QY 990 GTTTATCTCATCATTTATTTATCCGACTTTTCTATGTTGTCACAAATGAAAGATTAGGA 1049
DB 752 GTTTATATATTGATGAATATGTAAGTTGTAGTATTAATTTGTA-TTTTGAAGATTAGGA 810
QY 1050 ATTTTCACACCATTTTCAAGTGGAAATTAATAAATTTTATAGATGAACATACAGATTTGG 1109
DB 811 ATTGATAGAGATTTTAAAGAGTGAATAATAAAGATATTTGGATGAAA-ATATAGATATTGG 869
QY 1110 GTGGAACGAGATGAGCAAAATATTTCATGAGATGTTGTAACATGCTGCTTTAGCCTTTCCGTTA 1169

```

Db      870 GTTTGGGGGATGA--ACAATATGTTTGGATTGGGAATTTGGTTTGGCTTTTGGATTA 927
QY      1170 TTAAGGATCAATGGTATGAAGTTTCCCGAGATCCATTGGCGTGAATTAATTAATGAATTA 1229
Db      928 TTGTTGCTGATGGATATATGTTGTTTAAGATAAGATAAAAGATTTTGCAGGATTTG 987
QY      1230 GCTT 1233
Db      988 GTTT 991

RESULT 12
LOCUS   CF037118/18
DEFINITION OCG40H09.yg QCG Zea mays cDNA clone QCG40H09, mRNA sequence.
ACCESSION CF037118
VERSION   CF037118.1 GI:32932306
KEYWORDS EST.
SOURCE   Zea mays
ORGANISM Zea mays
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
          clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 543)
          Genoplatante.
          A major partnership french program in plant genomics
          Unpublished (2003)
          Contact: Genoplatante
          Genoplatante
          93, rue Henri Rochefort 91025 EVRY CEDEX France
          Tel: 33 1 69 47 54 00
          Fax: 33 1 69 47 54 10
          This sequence has been generated in the framework of the french
          plant genomics programme 'Genoplatante' (http://www.genoplatante.com)
          and http://genoplatante-info.infobiogen.fr.
          Location/Qualifiers
            1..543
              /organism="Zea mays"
              /mol_type="mRNA"
              /cultivar="F2"
              /db_xref="taxon:4577"
              /clone="QCG40H09"
              /tissue_type="embryo"
              /clone_lib="QCG"

FEATURES             source
    source
    1..543
      /organism="Zea mays" (indica cultivar-group)
      /mol_type="mRNA"
      /cultivar="F2"
      /db_xref="taxon:4577"
      /clone="QCG40H09"
      /tissue_type="embryo"
      /clone_lib="QCG"

ORIGIN
Query Match          7.0%; Score 195.8; DB 6; Length 543;
Best Local Similarity 61.1%; Pred. No. 2.4e-34;
Matches 317; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY      1505 CTCATCAAAATATTAGTAACACTGATTACCTAAGGTGGCTCTTTGAAGATTTCTACACCTG 1564
Db      538 CTCATCTTTTCGATCAATCAAGAATTCCTAGTCTTGGCAGTCAAGATTTCACTTCTC 479
QY      1565 CCAATCTATTATTATCGTGAAGAAATTAAGAGCTTTGAAGGTGGGTGGAGAGATAAGTT 1624
Db      478 TCACAGGTGTTACGGGATGAACCTCGCATCTTGATAGTTGGTGAAGAGACAACT 419
QY      1625 GGACCACTCAAGTTTGTCTAGGCAAAAGACCGCTACTGTTATTCTCTGTTGCTGCAAC 1684
Db      418 GGACCACTCAAAATTTGCTCGGCAGAAACTGACATATTGCTATCTGCTGCTGCTAC 359
QY      1685 ACTTTCGTCCTCCCAATATCAGATGCGCGTATTTCATGGGCCAAAATGGCATATTAAAC 1744
Db      358 CGTATTTCCTCTGAATTTGCTGACGCTCGCATTTTCATGGGCCAAAATGGTGCTCAC 299
QY      1745 TACAGTAGTTGATGACATTTTTTTGATATCGGTGGTACCAATGATGAAATGACCAACCTGAT 1804
Db      298 AACTGTGTTGATGACATCTCTTCGATGTTGGTGGATCAAAAGAAATTAGAAACCTGAT 239
QY      1805 TCAATGTTTGAAGAAATGGAATGTAGATGTGCAAGAGATTGTTGTTTCAGAGCATGTTG 1864

```

```

Db      238 AGCACTAGTTGAGAAATGGCATGCCACCATGCAGTTCTAGTTCTTATTCGGAACAGTGAA 179
QY      1865 GATTTTATTTTAGCANTTAAAGATGCAATCTGTTGGATTGGAGATGAAGCTTTTAAATG 1924
Db      178 AATAGTATTTTCTGCTATTATACAAAGATGAACCATCTTGGAGCAATGGCTTCTGCAGC 119
QY      1925 GCAAGCGCGCATGTAACTAGCCATGTTATTCAAACTTGGTTGGAACTAATGAATAGTAT 1984
Db      118 ACAAGCCGTGATCTTACAACCCACTAGTAGAAATATGGCTGGATTGTTTAAGATCTAT 59
QY      1985 GTTGAGAGAAGCTATATGGAACAAGAGATGCTTATGTGCC 2023
Db      58 GATGTCGAGGAGAAATGGCAGAGATGCCAATATGTACC 20

RESULT 13
LOCUS   CF197159
DEFINITION Ir64v11.5.E05 IR64-VI Oryza sativa (indica cultivar-group) cDNA clone IR64-VII.5E05 5', mRNA sequence.
ACCESSION CF197159
VERSION   CF197159.1 GI:33381816
KEYWORDS EST.
SOURCE   Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzaceae; Oryza.
          1 (bases 1 to 619)
          Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
          Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delsey,M. and
          Brugidou,C.
          Metabolism and Photosynthesis gene expression is affected in
          response to rice yellow mottle virus infection in Oryza sativa
          indica and Japonica cultivars
          Mol. Genet. Genomics (2003) In press
          Contact: Brugidou C
          IRD
          BP 64501, 34394 Montpellier cedex 5 France
          Tel: 33 4 67 41 67 39
          Fax: 33 4 67 41 61 81
          PCR Primers
          FORWARD: 5'AATTAAACCTCTCACTAAAGG3'
          BACKWARD: 5'CGGATATCACTCAGCATATG3'
          Plate: 1.5 row: E column: 5.
          Location/Qualifiers
            1..619
              /organism="Oryza sativa (indica cultivar-group)"
              /mol_type="mRNA"
              /cultivar="IR64"
              /db_xref="taxon:39946"
              /clone="IR64-VII.5E05"
              /tissue_type="leaves"
              /dev_stage="3 weeks-old plants"
              /lab_host="SOLR strain (Stratagene)"
              /clone_lib="IR64-VI"
              /notes="Vector: pBluescript SK (+/-); Site 1: Eco RI;
              Site 2: Xho I; Sample name: IR64-VI; Plant growth place:
              greenhouse; Soil conditions: standard; Sowing date:
              28/02/2000; Harvesting date: 17/03/2000; Stress date:
              15/03/2000; Cultivar: ir64; Description: IR64-RYM
              inoculated leaves cDNA library; experimental condition:
              rywv inoculation;"

ORIGIN
Query Match          6.8%; Score 190.4; DB 6; Length 619;
Best Local Similarity 56.1%; Pred. No. 4.4e-33;
Matches 345; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

QY      1393 TGAAGATGCTCTTAAGTTTCCCTATCAATACCGGTTTGAACGCATAAACACTAGACGA 1452
Db      2 TGAAGTATGCTCTTAAGTTTCCCTTCTACACCACCTGGACCGCTCTAGATCATCAGAGGA 61

```


/db_xref="taxon:4577"
/clone="QCG9g06"
/tissue_type="embryo"
/clone_lib="QCG"

ORIGIN

```

Query Match      6.6%; Score 185; DB 6; Length 515;
Best Local Similarity 61.4%; Pred. No. 7.6e-32;
Matches 315; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

Qy 1380 ATTACAAAGAGGTGGAAAATGCTTTAAAGTCCCTATCAATACCGGTTTAGAACGCATA 1439
Db 5 ATACTTGAAGAGGTAGAGTATGCCCTTAAGTTTCCCATTTATGCCGTAGTAGATCCTCTC 64

Qy 1440 AACACTAGACGAAATATACAGCTTTTACAAATGTAGACAATACAAGAAATCTGAAGACTACA 1499
Db 65 AACACAGAGAAGATGTGGAGCAATTTGGATGTAGGGGTTCTCAGACGCTAAGACAAAA 124

Qy 1500 TATCACTCATCAAAATATTAGTAACAATGATTACCTTAAGGTTGGCTGTTGAAGATTTCAC 1559
Db 125 TATCTGCCATGTCAATGTAAGTCA---AGACATCCCTTGGCTGTTGAAGATTTCGT 181

Qy 1560 ACCTGCCAATCTATTATCTGTGAAGAATTAAGGTCTTGAAGTGGGTGTAGAGAAAT 1619
Db 182 TTTTCTCAATCTATTATACAGGACGAACTACAGAAACATCATTAGTTGGGAGAAAGAAAT 241

Qy 1620 AAGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTATTCTCTGTGCT 1679
Db 242 AGGATGGACCAAGCTACAAATTTGTGGGCAAAAGGCTGGCAATTTGCTATCTCGTGTGCT 301

Qy 1680 GCAACACTTTCTGCTCCGAAATATCAGATGGCGTATTTTCATGGGCCAAAAATGGCATA 1739
Db 302 ACCACCATATCCCTCATGAATTTGCTGATGCTGCGTTGCATGTGCCAAAGCATCATG 361

Qy 1740 TTAACACAGTACTGTGATGACTTTTGTATCGGTGTAATATCGATGAATGACCAAC 1799
Db 362 CTCACGGTTGTAGTTGATGACTTCTTCGATGTTGGTATCAAAAGAGAAACAAGAAAAAT 421

Qy 1800 CTGATTCAATGTGTGAAAATGGAATGTAGATGTGCAAGAGGATGTTGTTTCAGAGCAT 1859
Db 422 CTCATCGAATTTAGTTGAACTGGGATGAGCAACAAAGTTGAGTTCTGTTCGGAGAAA 481

Qy 1860 GTTCGGATTTTATTATTAGCAATTAAGATGCA 1892
Db 482 GTAGAAATAGTTTCTATGCTGTCTATAATACA 514

```

Search completed: October 17, 2004, 03:24:33
Job time : 8685.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 23:38:29 ; Search time 1297.73 Seconds
(without alignments)
11005.915 Million cell updates/sec

Title: US-10-041-018-361

Perfect score: 2792

Sequence: 1 cccactcatctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3403857 seqs, 2557783690 residues

Total number of hits satisfying chosen parameters: 6807714

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

19: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

20: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|---------------------|---------------------|
| 1 | 2792 | 100.0 | 2792 | 16 | US-10-041-018-361 | Sequence 361, App |
| 2 | 2582.6 | 92.5 | 3117 | 16 | US-10-041-018-362 | Sequence 362, App |
| 3 | 1428.6 | 51.2 | 2594 | 16 | US-10-041-018-368 | Sequence 368, App |
| 4 | 809.4 | 29.0 | 2638 | 16 | US-10-041-018-367 | Sequence 367, App |
| 5 | 782.2 | 28.0 | 2658 | 16 | US-10-041-018-378 | Sequence 378, App |
| 6 | 782.2 | 28.0 | 2658 | 16 | US-10-041-018-381 | Sequence 381, App |
| 7 | 732.4 | 26.2 | 2506 | 16 | US-10-041-018-377 | Sequence 377, App |
| 8 | 561.2 | 20.1 | 2554 | 16 | US-10-041-018-3441 | Sequence 3441, App |
| 9 | 433.2 | 15.5 | 2086 | 16 | US-10-041-018-17004 | Sequence 17004, App |
| 10 | 392.4 | 14.1 | 2029 | 16 | US-10-041-018-5988 | Sequence 5988, App |
| 11 | 379.4 | 13.6 | 2223 | 16 | US-10-041-018-107 | Sequence 107, App |
| 12 | 356.4 | 12.8 | 2070 | 17 | US-10-041-018-19081 | Sequence 19081, App |
| 13 | 330 | 11.8 | 1559 | 16 | US-10-041-018-375 | Sequence 375, App |

| | | | | | | |
|----|-------|------|------|----|--------------------|--------------------|
| 14 | 315.6 | 11.3 | 3666 | 17 | US-10-041-018-361 | Sequence 72152, A |
| 15 | 288.8 | 10.3 | 1150 | 16 | US-10-041-018-361 | Sequence 729, App |
| 16 | 276.8 | 9.9 | 1008 | 16 | US-10-041-018-361 | Sequence 2457, App |
| 17 | 275 | 9.8 | 1615 | 17 | US-10-041-018-361 | Sequence 94395, A |
| 18 | 264.8 | 9.5 | 2193 | 17 | US-10-041-018-361 | Sequence 72156, A |
| 19 | 241.6 | 8.7 | 2313 | 17 | US-10-041-018-361 | Sequence 36043, A |
| 20 | 239.4 | 8.6 | 929 | 16 | US-10-041-018-361 | Sequence 51115, A |
| 21 | 238.4 | 8.2 | 1157 | 17 | US-10-041-018-361 | Sequence 19703, A |
| 22 | 198.4 | 7.1 | 711 | 17 | US-10-041-018-361 | Sequence 72153, A |
| 23 | 194.8 | 7.0 | 2861 | 9 | US-09-895-752-55 | Sequence 55, Appl |
| 24 | 194.8 | 7.0 | 2861 | 9 | US-09-887-586A-55 | Sequence 55, Appl |
| 25 | 194.8 | 7.0 | 2861 | 9 | US-09-903-012-55 | Sequence 55, Appl |
| 26 | 194.8 | 7.0 | 2861 | 10 | US-09-900-797-55 | Sequence 55, Appl |
| 27 | 194.8 | 7.0 | 2861 | 11 | US-09-893-820-55 | Sequence 55, Appl |
| 28 | 194.8 | 7.0 | 2861 | 13 | US-10-041-007-3 | Sequence 3, Appl1 |
| 29 | 194.8 | 7.0 | 2861 | 16 | US-10-041-018-365 | Sequence 365, App |
| 30 | 172.8 | 6.2 | 439 | 16 | US-10-041-018-373 | Sequence 373, App |
| 31 | 172 | 6.2 | 577 | 16 | US-10-259-194A-491 | Sequence 491, App |
| 32 | 155.4 | 5.6 | 575 | 17 | US-10-041-018-397 | Sequence 30695, A |
| 33 | 153.8 | 5.5 | 2700 | 9 | US-09-895-752-43 | Sequence 43, Appl |
| 34 | 153.8 | 5.5 | 2700 | 9 | US-09-887-586A-43 | Sequence 43, Appl |
| 35 | 153.8 | 5.5 | 2700 | 9 | US-09-903-012-43 | Sequence 43, Appl |
| 36 | 153.8 | 5.5 | 2700 | 10 | US-09-900-797-43 | Sequence 43, Appl |
| 37 | 153.8 | 5.5 | 2700 | 11 | US-09-893-820-43 | Sequence 43, Appl |
| 38 | 153.8 | 5.5 | 2700 | 16 | US-10-041-018-363 | Sequence 363, App |
| 39 | 133.8 | 4.8 | 2241 | 13 | US-10-041-007-38 | Sequence 38, Appl |
| 40 | 133.8 | 4.8 | 2388 | 13 | US-10-041-007-36 | Sequence 36, Appl |
| 41 | 133.8 | 4.8 | 2445 | 13 | US-10-041-007-34 | Sequence 34, Appl |
| 42 | 133.8 | 4.8 | 2622 | 13 | US-10-041-007-32 | Sequence 32, Appl |
| 43 | 133.8 | 4.8 | 2705 | 13 | US-10-041-007-1 | Sequence 1, Appl1 |
| 44 | 133.8 | 4.8 | 2705 | 16 | US-10-041-018-397 | Sequence 397, App |
| 45 | 130 | 4.7 | 339 | 11 | US-09-955-216-20 | Sequence 20, Appl |

ALIGNMENTS

RESULT 1
US-10-041-018-361
; Sequence 361, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; TITLE OF INVENTION: Biterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 361
; LENGTH: 2792
; TYPE: DNA
; ORGANISM: Stevia rebaudiana
US-10-041-018-361

| | | | | |
|-----------------------|-----------------|---|-----------|--------------|
| Query Match | 100.0%; | Score 2792; | DB 16; | Length 2792; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 2792; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | CCCACTCATCTTTATCAACCAATACCATCGTTCTGCCACCGGAAGCTGATAGCGACG | 60 | |
| Db | 1 | CCCACTCATCTTTATCAACCAATACCATCGTTCTGCCACCGGAAGCTGATAGCGACG | 60 | |
| Qy | 61 | GACCTGACGACGCTTTAACTCTGTGGACCACTAACAAAATCCGACCATCATGATCT | 120 | |
| Db | 61 | GACCTGACGACGCTTTAACTCTGTGGACCACTAACAAAATCCGACCATCATGATCT | 120 | |
| Qy | 121 | ACTCCGCTGACAGTTCATTTGCAATTTAGTAATCAGCAAACTAACATCATGATCT | 180 | |
| Db | 121 | ACTCCGCTGACAGTTCATTTGCAATTTAGTAATCAGCAAACTAACATCATGATCT | 180 | |

QY 181 TTTCACTATGATCGGTCCCTTTGTTAAACCAATCAAATCGAACCGCGGCTCTGTGAG 240
 Db 181 TTTCACTATGATCGGTCCCTTTGTTAAACCAATCAAATCGAACCGCGGCTCTGTGAG 240
 QY 241 CTATTCAVACAGCATCAACTTCCATGTTGGAGCAAACTTAATCCCACTAATCTGATCATTTG 300
 Db 241 CTATTCAVACAGCATCAACTTCCATGTTGGAGCAAACTTAATCCCACTAATCTGATCATTTG 300
 QY 301 AFACAAACCAAGAACGGATCCCAAAACAGTTTAAATAATGTAGAAATTTCTGTTTCTTCAAT 360
 Db 301 AFACAAACCAAGAACGGATCCCAAAACAGTTTAAATAATGTAGAAATTTCTGTTTCTTCAAT 360
 QY 361 ATGACACAGCATGGGTAGCCATGCTTCTCCAACTCACCCAAATCGCCTGTTTCC 420
 Db 361 ATGACACAGCATGGGTAGCCATGCTTCTCCAACTCACCCAAATCGCCTGTTTCC 420
 QY 421 CTGAGTGTCTCAATTTGGTTTAATAATCAATCAGCTTAATGATGGTTCATGGGTCCTTGTTA 480
 Db 421 CTGAGTGTCTCAATTTGGTTTAATAATCAATCAGCTTAATGATGGTTCATGGGTCCTTGTTA 480
 QY 481 ATCAGACTCATTAATCATATACCGGTTGCTTAAGATTTCTTAATCTTCAATAGCAT 540
 Db 481 ATCAGACTCATTAATCATATACCGGTTGCTTAAGATTTCTTAATCTTCAATAGCAT 540
 QY 541 GTATTGTTGCANTTAAAGATGGAATGTTGGGAAGATCAATAATAAAGGTCCTAAGTT 600
 Db 541 GTATTGTTGCANTTAAAGATGGAATGTTGGGAAGATCAATAATAAAGGTCCTAAGTT 600
 QY 601 TTATTGAGTCAATCTTGCTCAGCTACTGAAAGTCAACCATCTCCCATTTGGTTTG 660
 Db 601 TTATTGAGTCAATCTTGCTCAGCTACTGAAAGTCAACCATCTCCCATTTGGTTTG 660
 QY 661 ACATCATATTTCTGTTGCTTGTAGTATGCGAATAAATTTGACATAAACCCTTCCAA 720
 Db 661 ACATCATATTTCTGTTGCTTGTAGTATGCGAATAAATTTGACATAAACCCTTCCAA 720
 QY 721 AACAAACAGATTTAGTTGATGCTACATAAGAGGGAATTTGAGCAAAAAAGATGCCATT 780
 Db 721 AACAAACAGATTTAGTTGATGCTACATAAGAGGGAATTTGAGCAAAAAAGATGCCATT 780
 QY 781 CAAATGAGATGGAATGATCTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATT 840
 Db 781 CAAATGAGATGGAATGATCTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATT 840
 QY 841 GGAATATGTTGAAGAAATATCAGATGAAAAAATGGTTCTGTCTTCAACTCACCATCAGCAA 900
 Db 841 GGAATATGTTGAAGAAATATCAGATGAAAAAATGGTTCTGTCTTCAACTCACCATCAGCAA 900
 QY 901 CAGCTGCTCTTTCATTATCATCAAAATCCCTGGTTGCTTAATTTAAATTTCACTTT 960
 Db 901 CAGCTGCTCTTTCATTATCATCAAAATCCCTGGTTGCTTAATTTAAATTTCACTTT 960
 QY 961 TGGACAAGTTTGGTAATGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTT 1020
 Db 961 TGGACAAGTTTGGTAATGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTT 1020
 QY 1021 CTATGTTGACAAATTTGAAGATTTAGGAATTTCAACCATTTTCAAGTGGAAATTAATA 1080
 Db 1021 CTATGTTGACAAATTTGAAGATTTAGGAATTTCAACCATTTTCAAGTGGAAATTAATA 1080
 QY 1081 ATGTTTTAGATGAAACATACAGATGTTGGGTGGAAACGAGATGAGCAAAATTTATGAGATG 1140
 Db 1081 ATGTTTTAGATGAAACATACAGATGTTGGGTGGAAACGAGATGAGCAAAATTTATGAGATG 1140
 QY 1141 TTGTAAACATGTTGATGCTTTTGGTTTAAAGATCAATGGGTATGAAGTTTCCCGAG 1200
 Db 1141 TTGTAAACATGTTGATGCTTTTGGTTTAAAGATCAATGGGTATGAAGTTTCCCGAG 1200
 QY 1201 ATCCATTTGGTGAATTTCTAATGATTTAGTTTGAAGACGAATATGAGCTCTTGA 1260
 Db 1201 ATCCATTTGGTGAATTTCTAATGATTTAGTTTGAAGACGAATATGAGCTCTTGA 1260

QY 1261 CATATCATGCGTCACATATATATATACCAAGAGGATTTATCTTCTGAAAAACAAATCTTGA 1320
 Db 1261 CATATCATGCGTCACATATATATATACCAAGAGGATTTATCTTCTGAAAAACAAATCTTGA 1320
 QY 1321 AGTCAGCTGATTTCTCTCAAGAGATATATCCACTGATTTCAAAAGGCTTTCTTAATTTAA 1380
 Db 1321 AGTCAGCTGATTTCTCTCAAGAGATATATCCACTGATTTCAAAAGGCTTTCTTAATTTAA 1380
 QY 1381 TTCAACAGAGGTTGGAATGCTTTAAGTTCCCTATCAATACCGTTTGTAGAACGCATAA 1440
 Db 1381 TTCAACAGAGGTTGGAATGCTTTAAGTTCCCTATCAATACCGTTTGTAGAACGCATAA 1440
 QY 1441 ACACCTAGACGAAATATACAGCTTTTACAAATGTAGACAATAAAGAAATCTGAAAACTACAT 1500
 Db 1441 ACACCTAGACGAAATATACAGCTTTTACAAATGTAGACAATAAAGAAATCTGAAAACTACAT 1500
 QY 1501 ATCACTCATCAAAATATAGTAAACACTGATTAAGTTGCTTGAAGATTTCTTACA 1560
 Db 1501 ATCACTCATCAAAATATAGTAAACACTGATTAAGTTGCTTGAAGATTTCTTACA 1560
 QY 1561 CCTGCCAATCTATTTATCGTGAAGAAATTAAGAGGCTTCTGAAAGGTTGGGTAGAGATA 1620
 Db 1561 CCTGCCAATCTATTTATCGTGAAGAAATTAAGAGGCTTCTGAAAGGTTGGGTAGAGATA 1620
 QY 1621 AGTTGGACCGAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTGTTATTTCTCTGTTGCTG 1680
 Db 1621 AGTTGGACCGAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTGTTATTTCTCTGTTGCTG 1680
 QY 1681 CAACTCTTCTGCTCCGAAATTTATCAGATCGCGTATTTCAATGGGCAAAATTTGCAATAT 1740
 Db 1681 CAACTCTTCTGCTCCGAAATTTATCAGATCGCGTATTTCAATGGGCAAAATTTGCAATAT 1740
 QY 1741 TAACTACAGTACTTGTATGATCTTTTTTATATCGTGGTGAACAATCGATGAATTTGACCAAC 1800
 Db 1741 TAACTACAGTACTTGTATGATCTTTTTTATATCGTGGTGAACAATCGATGAATTTGACCAAC 1800
 QY 1801 TGATTCATTTGTTGAAAAATGGAATGTAGATGTGCAAGGATTTGTTTCTGAGAGCATG 1860
 Db 1801 TGATTCATTTGTTGAAAAATGGAATGTAGATGTGCAAGGATTTGTTTCTGAGAGCATG 1860
 QY 1861 TTCCGATTTTATTTTGTAGCAATTTAAAGATGCAATCTGTTGGATTTGGAGATGAGCTTTTA 1920
 Db 1861 TTCCGATTTTATTTTGTAGCAATTTAAAGATGCAATCTGTTGGATTTGGAGATGAGCTTTTA 1920
 QY 1921 AATGGCAAGCGCGCATGTAACTAGCCATGTTTATTCAAACTTTGGTTGGAACCTAATGAATA 1980
 Db 1921 AATGGCAAGCGCGCATGTAACTAGCCATGTTTATTCAAACTTTGGTTGGAACCTAATGAATA 1980
 QY 1981 GTATGTTGAGAGAAGCTATATGCAAGAGATGCTTTATGTCACCAATTAATGAATATA 2040
 Db 1981 GTATGTTGAGAGAAGCTATATGCAAGAGATGCTTTATGTCACCAATTAATGAATATA 2040
 QY 2041 TGGAAAAACCTTTACGTGTCTATTTGCAATTTAGGCGCGATTTCTCAGCCGCTATTACTTTG 2100
 Db 2041 TGGAAAAACCTTTACGTGTCTATTTGCAATTTAGGCGCGATTTCTCAGCCGCTATTACTTTG 2100
 QY 2101 TGGGGCCCCAAATTTATCAGAGGAGATTTGTTGAAAGCTCTGAAATATCATATCTATTTAAGC 2160
 Db 2101 TGGGGCCCCAAATTTATCAGAGGAGATTTGTTGAAAGCTCTGAAATATCATATCTATTTAAGC 2160
 QY 2161 TTAATGAGCAACGAGGCTGACCTTTAAACGATATCCATAGCTTTCAAGAGGGAATTTAAGC 2220
 Db 2161 TTAATGAGCAACGAGGCTGACCTTTAAACGATATCCATAGCTTTCAAGAGGGAATTTAAGC 2220
 QY 2221 AAGGCAAAATTAACCGGTTAGCATTTGCAATTCAGTAACCGGAAAGTGGAAAGTGGAAAG 2280
 Db 2221 AAGGCAAAATTAACCGGTTAGCATTTGCAATTCAGTAACCGGAAAGTGGAAAGTGGAAAG 2280
 QY 2281 AAGAGGTTTGTGAGGAGATGATCATGATTAATAAACAAGAGGAAAGAAATTAATGAAT 2340
 Db 2281 AAGAGGTTTGTGAGGAGATGATCATGATTAATAAACAAGAGGAAAGAAATTAATGAAT 2340
 QY 2341 TAAATTTTGAAGAAATGATGATTTGTTCTTAGAGCTTTGTAAGATGATTTTGGAAACA 2400

```
Db 2341 TAATTTTGAAGAAATGCTAGCATGTTCTTAGAGCTTTGTAAGATGCATTTTGGAA 2400
Qy 2401 TGTGTCACGCTGTTGAATTTTTTTTACGCAACGATGACGGGTTTACTGGAACACGATTC 2460
Db 2401 TGTGTCACGCTGTTGAATTTTTTTTACGCAACGATGACGGGTTTACTGGAACACGATTC 2460
Qy 2461 TTGATCTGTGAAGGACATCAITTTACAAACCGGTTGGTCTGTGTAATGAAATGAAGAAC 2520
Db 2461 TTGATCTGTGAAGGACATCAITTTACAAACCGGTTGGTCTGTGTAATGAAATGAAGAAC 2520
Qy 2521 AAAGGTAATGGCCTTTAAAGATGATATATATATATATATATATATATATATATATATAT 2580
Db 2521 AAAGGTAATGGCCTTTAAAGATGATATATATATATATATATATATATATATATATATAT 2580
Qy 2581 TAGTTGTATATAAAGGTTGTAGTTGCTATAATAAAGGTAATAGTAATCAATAGAAAGCT 2640
Db 2581 TAGTTGTATATAAAGGTTGTAGTTGCTATAATAAAGGTAATAGTAATCAATAGAAAGCT 2640
Qy 2641 TAAAGTTATTAAGTTTTTCCCTCGTGTACACACGGGTGAGGTTCTTGTAAAGCAGTTTAT 2700
Db 2641 TAAAGTTATTAAGTTTTTCCCTCGTGTACACACGGGTGAGGTTCTTGTAAAGCAGTTTAT 2700
Qy 2701 TTTATGTTTGTACACACCTTGGTGTGTTGTTGTTGTTTATTTCTTAAGAGGCTTTCACT 2760
Db 2701 TTTATGTTTGTACACACCTTGGTGTGTTGTTGTTGTTTATTTCTTAAGAGGCTTTCACT 2760
Qy 2761 TTGTTCTTAAAGGTTTTTAAAGGTTTTTAAAGGTTTTTAAAGGTTTTTAAAGGTTTTTAA 2792
Db 2761 TTGTTCTTAAAGGTTTTTAAAGGTTTTTAAAGGTTTTTAAAGGTTTTTAAAGGTTTTTAA 2792

RESULT 2
US-10-041-018-362
; Sequence 362, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 362
; LENGTH: 3117
; TYPE: DNA
; ORGANISM: Stevia rebaudiana
US-10-041-018-362

Query Match 92.5%; Score 2582.6; DB 16; Length 3117;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2603; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 140 TTTTCAATTTAGTAAATCAGCAAACTAAACATCAATCAATCTTCACTATCGATCGGTC 199
Db 466 TTGGTGTGAGTAATCAGCAAACTAAACATCAATCAATCTTCACTATCGATCGGTC 525
Qy 200 CCCTTTGTTAAACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 259
Db 526 CCCTTTGTTAAACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 585
Qy 260 TTCACTGTTGGACAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAA 319
Db 586 TTCACTGTTGGACAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAA 645
Qy 320 CCAAAAAACAGTTTAAAGGTTAGAAATTTCTGTTTCTTCACTATCACTAGACAGATGGTAG 379
Db 646 CCAAAAAACAGTTTAAAGGTTAGAAATTTCTGTTTCTTCACTATCACTAGACAGATGGTAG 705
```

```
Qy 380 CATGTCCTCTCTCCAACTCACCACAAATCGCCTTGTCTTCCCTGAGTGTCTCAATGGTT 439
Db 706 CATGTCCTCTCTCCAACTCACCACAAATCGCCTTGTCTTCCCTGAGTGTCTCAATGGTT 765
Qy 440 AATTAATATACAGCTTAATGATGGTTCAATGAGGTTCTGTTTAAATCACTCAATATCAATA 499
Db 766 AATTAATATACAGCTTAATGATGGTTCAATGAGGTTCTGTTTAAATCACTCAATATCAATA 825
Qy 500 TCACCGGTTGCTTAAAGATCTCTATCTTCAACATTTAGCATGTATTTGTCATTTAAAG 559
Db 826 TCACCGGTTGCTTAAAGATCTCTATCTTCAACATTTAGCATGTATTTGTCATTTAAAG 885
Qy 560 ATGGAATGTTGGGAGATCAATTAATAAAGGTTCTAAGTTTATTTAGTCAATCTTGC 619
Db 886 ATGGAATGTTGGGAGATCAATTAATAAAGGTTCTAAGTTTATTTAGTCAATCTTGC 945
Qy 620 TTCACTACTGAAAAAGTCAACCATCTCCATCTGGTTTGTGACATCATATTTCTGGTT 679
Db 946 TTCACTACTGAAAAAGTCAACCATCTCCATCTGGTTTGTGATATCATATTTCTGGTT 1005
Qy 680 GTTTCAGTATGCGAAAAAATTTGGACATTAACCTCTCTTCAAAACAAACAGATTTTGT 739
Db 1006 GTTTCAGTATGCGAAAAAATTTGGACATTAACCTCTCTTCAAAACAAACAGATTTTGT 1065
Qy 740 GATGCTACATAAGAGGAAATGGAGCAAAAGATGCCATTTCAATGAGATGGATGATA 799
Db 1066 GATGCTACATAAGAGGAAATGGAGCAAAAGATGCCATTTCAATGAGATGGATGATA 1125
Qy 800 CTGCGCTATATCTCTGAAGGACTCGGTAATTTATATGATTTGGAATATGGTGAAGAAATA 859
Db 1126 CTGCGCTATATCTCTGAAGGACTCGGTAATTTATATGATTTGGAATATGGTGAAGAAATA 1185
Qy 860 TCAGATGAAAAAATGTTCTGTTTCAACTCAACATCAGCAACAGCTGCTGCTTCAATTA 919
Db 1186 TCAGATGAAAAAATGTTCTGTTTCAACTCAACATCAGCAACAGCTGCTGCTTCAATTA 1245
Qy 920 TCATCAAAATCCTGTTGCTTAATTTAAATTTCACTTTTGGACAGTTTGGTAATGC 979
Db 1246 TCATCAAAATCCTGTTGCTTAATTTAAATTTCACTTTTGGACAGTTTGGTAATGC 1305
Qy 980 AGTCCCAACAGTTTATCTCATGATTTATTTATCGACTTTCTATGTTGACACAAATGA 1039
Db 1306 AGTCCCAACAGTTTATCTCTGATTTATATATCGGCTTTCTATGTTGACACAAATGA 1365
Qy 1040 AAGATTTAGGAAATTTCAACCATTTTCAAGTGGAAATTTAAATTTTATGATGAACATA 1099
Db 1366 AAGATTTAGGAAATTTCAACCATTTTCAAGTGGAAATTTAAATTTTATGATGAACATA 1425
Qy 1100 CAGATGTTGGTGGAAACGAGATGAGCAATTTTCAATGATGTTTGAATGCTGCTTAC 1159
Db 1426 CAGATGTTGGTGGAAACGAGATGAGCAATTTTCAATGATGTTTGAATGCTGCTTAC 1485
Qy 1160 CTTTTCGGTTATTAAGGATCAATGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTA 1219
Db 1486 CTTTTCGGTTATTAAGGATCCACGGTATTAAGTCTCCCGAGATCAATTTGGCTGAAATTA 1545
Qy 1220 TAATGAATTTAGCTTTGAAAGACGAATATGAGCTCTTTGAAACATATCATGCTGACAT 1279
Db 1546 TAATGAATTTAGCTTTTCAAGACGAATATGAGCTCTTTGAAACATATCATGCTGACAT 1605
Qy 1280 ATTATACCAAGAGGATTTATCTTCTGGAACAAATCTTGAAGTCACTGATTTCTTCA 1339
Db 1606 ATTATACCAAGAGGATTTATCTTCTGGAACAAATCTTGAAGTCACTGATTTCTTCA 1665
Qy 1340 AGAGATAATATCACTGATTTCAACAGGCTTTCTTAAATTAATTTCAAAAGAGGTTG 1399
Db 1666 AGAGATAATATCACTGATTTCAACAGGCTTTCTTAAATTAATTTCAAAAGAGGTTG 1725
Qy 1400 TGCTTTAAGTTCCTTATCAATACCGGTTTGAACGCAATAAACACTAGACGAATATACA 1459
Db 1726 TGCTTTAAGTTCCTTATCAATACCGGTTTGAACGCAATAAACACTAGACGAATATACA 1785
Qy 1460 GCTTTACAATGTAGACATACAAGATTTCTGAAAACTATACATATCATCTCAATATTAG 1519
```


Db 655 TAAACCAACACATTTGAGTGTGATGTTACATGAGAGAAATGAGCTAAGAGATGTC 714
QY 778 ATTCAAATGAGATGATGGATCTTGGCGTATATCTCTGAAGGACTCGGTAAATTAATG 837
Db 715 ATTCAAATGAGGAGGAAACATCTTGGCATATATCTCAAGAGGACTTGGAAATTTAAATG 774
QY 838 ATTGGAATATGCTGTAAGAAATATCAGATCAAAATGGTTCGTCTTTCAACTCAACATCAG 897
Db 775 ACTGGAACATGGTGTGATGAATATCAATGAAGATGGTTCCTTTTCAACTCACCTCAG 834
QY 898 CAACAGCTGCTGCTTTCATTAATCAATCAAAATCCCTGGTGTCTTAATTAATTTAAATTCAC 957
Db 835 CAACAGCTCTGCTTCTTAATCATCATCAAAATGCTGGTGTCTTCACTATCACTTCACTTCA 894
QY 958 TTTTGGCAAGTTGGTGAATGCAAGTCCCAACAGTTTATCCCATCATGATTTATTTATCCGAC 1017
Db 895 TCCTTGGCAAAATTTGGAATATGCAAGTCCCAACAGTTTATCCCTATTAATTTATATGAGGC 954
QY 1018 TTTCTATGCTTGACACAAATGAAAGATTAGGAATTTTCAACCAATTTCAAGTGGAAATTA 1077
Db 955 TTTCCATGCTTGACACACTTGAAGATTGGGAATCAAGAGCAATTTCAATGTTGGAAATCC 1014
QY 1078 AAAATGTTTATGAGTAAACATACATAGATGTTGGTGGAAACGAGAGCAAAATATTTCATGG 1137
Db 1015 AAAATGTTTATGAGTAAACATATAGATGTTGGTGCAGGGGATGTCAAATATTTCATGG 1074
QY 1138 ATGTTGTAACATGCTGCTTACCTTTCGTTTATTAAGGATCAATGGGTATGAAGTTTCCC 1197
Db 1075 ATGTTGTCACATGCTGCTTACCTTTCGTTTATTAAGGATCAATGGGTATGAAGTCTCTT 1134
QY 1198 CAGATCCATGCTGAAATTTACTTA-----TGAATTAGCTT 1233
Db 1135 CGGATCCATGCTGCTTAAATCCAAAGAGGGGATTAATGAATTCACCTGAAAGCTT 1194
QY 1234 TGAAGACGAATATGAGCTCTTGAAACATATCATGCTCTCATATATATATACCAAGAGG 1293
Db 1195 TCAAGAGCTATACATCTCTTGAAGTTTATAGGCAATCAAGATCATATATCAAGAGG 1254
QY 1294 ATTTATCTCTGGAAGAAACAAATCTGAAGTCAAGTCAAGTATTTCTTCAAGAGTAAATATCCA 1353
Db 1255 AGTTAGCTTTTAGAGAACAAATCTGA-----CAAGTTATCTCC 1293
QY 1354 CTGATTCAAACAGGCTTCTTAAATTAATTCAAAGAGGTGGAAATGCTCTTAAAGTTCC 1413
Db 1294 CTTCAGTAAACAGCTCTTAATATATTTCTCAAGAGGTGAGTGTCTTAAAGTTCC 1353
QY 1414 CTATCAATACCGGTTTGAAGCGATATAACATAGACGAAATATACAGCTTTTCAATGTAG 1473
Db 1354 CTTTAAATGGCAGCTAGAACGCTATGCTCCATAGAAAGAAACATAGAGCATTAACAATTTAA 1413
QY 1474 ACAAATACAGAAATCTGAAGAACTACATCTCATCAAAATATTAGTAACTCATTTACC 1533
Db 1414 ACCATCAAGAAATCTTGAAGAACTACATATAGCTCATCAAAATATTAGCAACAAAGATTACC 1473
QY 1534 TAAGTTGGCTGTTGAAGATTCTTACACCTGCCAATCTAATTTATCGTGAAGAAATTTAAAG 1593
Db 1474 TAAAGTTAGCTGTGCAAGATTTCAATGAATGCCAATCTATCTATTGTGAAGAACTAAAG 1533
QY 1594 GTCTTGAAGAGTGGTGTGAGAAATTAAGTTGGACCCAGCTCAAGTTTGTGTAGGCAAAAGA 1653
Db 1534 ATCTTGAAGAGTGGTGTGAGAAACAGATTAGACAAGCTCAAAATTTGTAGACAGAAGA 1593
QY 1654 CCGCTACTGTTATTTCTCTGCTGTCGCAACACTTTCGTCCTCCGAAATTTACAGATGGC 1713
Db 1594 CAGCTACTGTTACTTTTCTGCTGCTTCAATTTCTTCTTCCCTGATTATTCAGATGCC 1653
QY 1714 GTATTTTCATGGCCAAATATGGCATATTAACTACAGTATTGTAGTACTTTTTCATATCG 1773
Db 1654 GAAATTCATGGCCAAAGTAGCATCTCTACTACAGTAATTTGAAGACTTTTTCATGTTG 1713
QY 1774 GTGTACATCGATGAATGACCAACCTGATTCATGTTGTTGAAATTTGAATGTAGATG 1833

Db 1714 GAGGATCTATGATGAAGTGTGGAACCTTTGTTTCATCATTTGAAATAATGGAATGAAAG 1773
QY 1834 TCGACAACGATTTGTTTTCAGAGCAATGTTTCGATTTTATTTTACGATTTAAAGATGCAA 1893
Db 1774 TTGAAATGATTTGCTGTTTCAGAGGAAGTTGGCGTTTATTTTATTTTACACTTAAAGATGCGAG 1833
QY 1894 TCTGTTGATTTGGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAATTAATCCCATGTTA 1953
Db 1834 TCTGTTGATTTGGAGACAAAGCTTTTAAATTCAGAAACGCAATATAAAGCCCATGTTA 1893
QY 1954 TTCAAACTTGGTGGAACTAATAGTATGTTTGGAGAAAGCTATATGAGCAAGAGATG 2013
Db 1894 TTGAAATTTGTTGGATTTGTTGTAAGAGTATGTTTGGAGAAAGCTATATGGCGGAAAGATG 1953
QY 2014 CTTATGTCGAACATTAATGAATATATGGAAGAACGTTTACGTGTTCATTTGATTTAGGCC 2073
Db 1954 GTTCAATTTCCAACATAAATGAATATATGGAATAATGTTTACGAGGTAATTTGTCGAGGAAGTTGTTTCA 2013
QY 2074 CGATTGTCAGCCGCTATTTTACCTTTGTTGGGCGCCCAATTTATCAGAGGAGATGTTGTA 2133
Db 2014 CGATTGTCCTCCCTACCTTTTACCTTTTGGGGTAAATTTGTCGAGGAAGTTGTTTCA 2073
QY 2134 GCTCTGAATATCAATACTATTTAAGCTAATGAGCACGCGGCTCGACTTCTTAAACGATA 2193
Db 2074 GCTCGAGTATCACAGCTTTTATGAGGTTATGAGCACTCAGGCTCGTCTCATGATGATA 2133
QY 2194 TCCATAGCTTCAAGAGGAAATTTAAGAGGCAATTAATAACGCGGTAGCATTTGCAATTTGA 2253
Db 2134 TCCATAGCTTCAAGAGGAAAGAGGCGGGGAAATTTGAATGCTGTGGCATTTGATACATGA 2193
QY 2254 GTAAACGGAAGAAAGTGGGAAAGTGAAGAAAGAGGTTGTGGAGGAGATGATGATGATTA 2313
Db 2194 GTGATGGAAGAGTGGGAGTGTGGAAGAGAGGTTGTAGAGGAGATGAAGATTTTGACAA 2253
QY 2314 AAAACAAGAGAAAGAAATTAATGAAATTAATTTTGAAGAAATTTGTTAGCAATTTGTTCTTA 2373
Db 2254 AAAGTCAAGAGGAAAGAGATGATGAAATTTAGTTTGGAAACAAAGGAAGTGTGTTTCCA 2313
QY 2374 GAGCTTGAAGATGCAATTTTGGAAACATGTCAGGTGTGAAATTTTATTTTATGCAAAAG 2433
Db 2314 GAGTATGCAAGATGTAATTTTGGAAACATGTCAAATGTTTGAATCTTTTATGCGACAG 2373
QY 2434 ATGAGGCTTTTACGGAACACGATTTCTGATATCTGTGAAGGAGATCATTTTACACCGT 2493
Db 2374 ATGATGGTTCCTGGAATATGCCATTTCTGATGTTGCAAGGAAATATATATGAACCTG 2433
QY 2494 TGGTCTTGTGAATGAAATGGAAGAACAAAGGTAATTTGGCCTTTTAAAGATGATAATAAT 2553
Db 2434 TGTCCCATGAGTTGATATAATCAATTTGTTATATACCAATTCGTATTATATTTATGATCTC 2493
QY 2554 ATCATGCTTTGACGGGGTACTGTTGATGTTGTAATTAAGGTTGTAGTTGTTGTTATATA 2613
Db 2494 AAAAATGAAACAAAGCATACTAGTATTCTAGTACTCTCATGTTTATTTATTTTATTTT 2553
QY 2614 AAGGTAATAGTAAATCAAT 2632
Db 2554 AATATAGTTTGTAGCTAT 2572

RESULT 4

US-10-041-018-367
; Sequence 367, Application US/10041018
; Publication No. US2004007232A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413

| | | | | | | | | | |
|--|------|---|------|--|--|--|--|--|--|
| ; SOFTWARE: PatentIn version 3.1 | | | | | | | | | |
| ; SEQ ID NO 367 | | | | | | | | | |
| ; LENGTH: 2638 | | | | | | | | | |
| ; TYPE: DNA | | | | | | | | | |
| ; ORGANISM: Cucumis sativus | | | | | | | | | |
| US-10-041-018-367 | | | | | | | | | |
| Query Match 29.0%; Score 809.4; DB 16; Length 2638; | | | | | | | | | |
| Best Local Similarity 62.0%; Pred. No. 6.7e-180; | | | | | | | | | |
| Matches 1410; Conservative 0; Mismatches 811; Indels 54; Gaps 6; | | | | | | | | | |
| QY | 267 | GGTGACAACTAAATCCCACTAATCTGATCAATGTATACCAACCAAGAAAGCGATCCAAAA | 326 | | | | | | |
| DB | 159 | GCACAAAACTAAACTGGAGCTTTCGCTTTGAAGAAACAAAGAAAGATAAAAAA | 218 | | | | | | |
| QY | 327 | CAGTTTAAAAATGTGAATTTCTGTTCTTCATATGACACAGCATGGGTAGCCATGGTC | 386 | | | | | | |
| DB | 219 | TTGTTCAAAAATGTTGAACCTTTCATTTCTGCATATGATACGATGGGTGGCAATGGTC | 278 | | | | | | |
| QY | 387 | CCTTCTCCAAATCCACCAAAATCGCTTTTCCCTGAGTGTCTCAATTTGGTTAAATTAAT | 446 | | | | | | |
| DB | 279 | CCTTCTCCAAATCTCTTAATAAACCCTTTTTCTGAGTGATTAACCTGGGTATTAGAT | 338 | | | | | | |
| QY | 447 | AATCAGCTTTAATGATGGTTTCATGGGGTCTTGTTAATCACATCATATAATCATACCCG | 506 | | | | | | |
| DB | 339 | CATCAAAACCCCTGATGGGTCAATGGG-----CATACTCCATGACCATCAGTTG | 386 | | | | | | |
| QY | 507 | TTGCTTAAAGATCTCTATCTCAACATTAGCATGTATTTGTCATTAATAAGATGAAT | 566 | | | | | | |
| DB | 387 | GTGATGAAGCCACTCTCTATTCACATTAGCATGTGTCTTTACTCTTAAGCGATGGAT | 446 | | | | | | |
| QY | 567 | GTGCGGAAGATCAATAATAAGCTCTAAGTTTTATTGAGTCAAAATCTTCTTCAGCT | 626 | | | | | | |
| DB | 447 | ATCGGTGATGATCATATGAGCAAGGCCCTTAGTTTTATCAAGTCTAATATAGCTTCAGCT | 506 | | | | | | |
| QY | 627 | ACTGAAAAAGTCAACCATCTCCATTTGGTTTTGACATCATATTTCCCTGGTTTGGTTAG | 686 | | | | | | |
| DB | 507 | ACTGATGAGAACTCAACGTTCTCTGTGGGATTTGACATAATTTTCCCTGGTATGATTGAG | 566 | | | | | | |
| QY | 687 | TATGCGAAAAACTTGACATATAACCTCTTCAAAACAACAGATTTTATGTTGATGCTA | 746 | | | | | | |
| DB | 567 | TATGCTAAAGACTGAATTTGATTTGATCTACCTTTGGCATCAATGAATTTGGATGCTTTGGTT | 626 | | | | | | |
| QY | 747 | CATAAGAGGAATTGAGCAAAAAAGAT-----GCCATTCAAATCAGATGGATGATAC | 800 | | | | | | |
| DB | 627 | CAAAAGAAAGTTGGAGCTTAGAGCTGTGTAGCACTCTGAGAGGAAAGCCTAT | 686 | | | | | | |
| QY | 801 | TTGGCGTATATCTCTGAAGGACTCGGTAATTTATATGATTTGGAATATGTTGAAGAATAT | 860 | | | | | | |
| DB | 687 | TTAGCGTATGTTTCAGAAAGAAATTGGAATAATTACAGGACTGGGAATGGTCAATGCGATAT | 746 | | | | | | |
| QY | 861 | CAGATGAATAATGGTTCTGTCTTTCAACTCACCATCAGCAACAGCTGCTGTTTCATTAAT | 920 | | | | | | |
| DB | 747 | CAAGGAAGAACCGGATCACTGTTTAGTTCTCCATCCACCACGCGCAGTGGCTTTATGCGAC | 806 | | | | | | |
| QY | 921 | CATCAAAATCCTGGTTGTCTTAATTAATTTAAATTTCACTTTTGCAACAAGTTTGGTAATGCA | 980 | | | | | | |
| DB | 807 | AGAAATGATGAGCTGTTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 866 | | | | | | |
| QY | 981 | GTCCCAACAGTTTATCCTCATGATTTATTTAATCCGACTTTCTATGTTGACACAAATGAA | 1040 | | | | | | |
| DB | 867 | GTTCGCGCAATATATCTCTTGATATATATGCTCGTTTACACATGGTTGATAGCCTTCAA | 926 | | | | | | |
| QY | 1041 | AGATTAGGAATTTCAACCAATTTGAGTGGAAATTAATAATTTTATGATGAACATATC | 1100 | | | | | | |
| DB | 927 | AACTGGGGATGATGGCCATTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT | 986 | | | | | | |
| QY | 1101 | AGATGTTGGGTGAAACGAGATGAGCAAAATATTCAATGAGTGTGTAACATGCTGTTAGCC | 1160 | | | | | | |
| DB | 987 | AGCTGTTGGATGCAAGGGAGGAAACATATCTTAGATGCTTCAACTTGTGCAATGGCC | 1046 | | | | | | |
| QY | 1161 | TTTCGGTTATTAAGGATCAATGGGTATGAAGTTTCCCAGATCCATTTGGCTGAAATTAAT | 1220 | | | | | | |
| DB | 1047 | TTCCGGATGTTACGTTTCAAGGATATGATGTTTCTTCAGATCAATTTGACTCAATTTCTCA | 1106 | | | | | | |
| QY | 1221 | AA-----TGAATTAGCTTTGAAGACGATATATGACCTCTTGAA | 1259 | | | | | | |
| DB | 1107 | GAAGTCTCTTTTCAAAATGCTCGGAGGACATTTAAAGACTTTAGTGCCTCACTAGAG | 1166 | | | | | | |
| QY | 1260 | ACATATCATGCTGCACA---TATATTTATACCAAGAGGATTTATCTTCTGGAACAAATC | 1316 | | | | | | |
| DB | 1167 | TTATTTAAGGCTCCAGATTTATCAATTTATCCGGATGAGTTTATCTCGAAATATAAAC | 1226 | | | | | | |
| QY | 1317 | TTGAAGTCAGCTGATTTCTCTCAAGAGATATATCCACTG-----ATTCAACAGG | 1367 | | | | | | |
| DB | 1227 | TCTTGGACTAGTTCGTTTCTGTAATCATGATTTATCTAGTGGTTCAGTTCAATTTCTGATA | 1286 | | | | | | |
| QY | 1368 | CTTTCTAAATTAATTTCAAAAGAGGTGAAATGCTTTAACTTCCCTATCAATCACTCGGT | 1427 | | | | | | |
| DB | 1287 | ACTGAGAGACTCGTGAACACAGAGGAGTTAATGCTTCGAGTTCCCTATATATTTCACT | 1346 | | | | | | |
| QY | 1428 | TTAGAACGCATAAAACACTAGACGAATATACAGCTTTTACAATGTAGACAAATCAGAAAT | 1487 | | | | | | |
| DB | 1347 | CTGGAACGCTTATCAATAAAGCGACACTGGAAAGTTACAGTGGAGACATTTGTGAGGAT | 1406 | | | | | | |
| QY | 1488 | CTGAAAACTACATATCATCATCAATATTAGTAACACTGATTACTTAAGTTTGGCTGTT | 1547 | | | | | | |
| DB | 1407 | TCAAAAACAGCATATGCTGCTTAAATTTTGTTCATCAAGATTTCTTGGAACTTGCTGTA | 1466 | | | | | | |
| QY | 1548 | GAGATTTTCTACACCTGCCAATCTATTTATCTGGAAGATTTAAAGGTCTTGAAGGTGG | 1607 | | | | | | |
| DB | 1467 | GAGATTTTCAATCCCTGCAAGCATATCCCAAGAACTGAAAGAGCTTGAANAATGG | 1526 | | | | | | |
| QY | 1608 | GTGTAGAGAAATPAAGTTGGACCAAGCTCAAGTTTGTGAGCAAAAGACCCCTACTGTTAT | 1667 | | | | | | |
| DB | 1527 | GTCTGCAAAACAAATTTGGCAAGTTGAAATTTGCGAGACAGAAAGTTAGCTACTGCTAT | 1586 | | | | | | |
| QY | 1668 | TTCTCTGTGTCGCAACACTTTCTGCTCCGAAATTTACAGATGCGGATTTTCATGGGCC | 1727 | | | | | | |
| DB | 1587 | TTTTCTGTGTCGACGACCCCTAACCTCTCCAGAACTTTGTGATGCCGCTTATCATGGCA | 1646 | | | | | | |
| QY | 1728 | AAAAATGGCATTTAACTACAGTAGTTGATGATCTTTTGTGATATCGGTGGTACAATCGAT | 1787 | | | | | | |
| DB | 1647 | AAAAATGGGTACTCAACCTGTTGTGATGATTTCTTTGATGCTTGGAGAACTGGAAG | 1706 | | | | | | |
| QY | 1788 | GAAATGACCAACTGATTCATTTGTGAAAAAAGAAATGTAGATGTGCAAGAGATTTGT | 1847 | | | | | | |
| DB | 1707 | GAAATGGTAAACCTTATACAAATTTGGTGGAAAAAGTGGATGCGAGTGGGAAACGGGTAC | 1766 | | | | | | |
| QY | 1848 | TGTTGAGCATGTTCCGATTTTATTTTAGCATTAAGATGCAATCTGTTGGATTGGA | 1907 | | | | | | |
| DB | 1767 | TGTTCCAGGAGTTGAGATTTATTTCTTGCACTTCATAGCACAATTTGTGAAATAGGA | 1826 | | | | | | |
| QY | 1908 | GATGAAGCTTTTAAATGGCAAGCGCGCATGTAACCTAGCATGTTTATTTCAAACTTGGTTG | 1967 | | | | | | |
| DB | 1827 | AAAAAGCTTTTACCTTGGCAAGACGCGCTGATGAGGAAATGTTATCGATATTTGGTTG | 1886 | | | | | | |
| QY | 1968 | GAACTAATGAATGATGTTGAGAGAGCTATATGAGCAAGAGATCTTATGTGCAACA | 2027 | | | | | | |
| DB | 1887 | GCTTTGCTCGCTCAATGAGGAAGAAAGCTGAAATGTTGAAAAATAAGGTAGTGCATCA | 1946 | | | | | | |
| QY | 2028 | TTAATGAATATATGAAACGCTTACCTGCTCATTTGCTATTTAGCCCGGATTTGCAAGCG | 2087 | | | | | | |
| DB | 1947 | TTGGATGATACATGAAAAATGGCTATGATGATCATTTTGTCTTTTGGGACCTATAGTCTTCCA | 2006 | | | | | | |
| QY | 2088 | GCTATTATCTTTTGGGGGCCAAATTTATCAGAGGAGATTTGTTGAAAGCTCTGAATATCAT | 2147 | | | | | | |
| DB | 2007 | ACGCTCTACTTTTGTGACCTTAAGCTTCCAGAGAAATTTGTTGAAATTTGTGAATACAAG | 2066 | | | | | | |
| QY | 2148 | AATCTAATTAAGTAATGAGCGAGGCTGAGCTTCTTAAACGATATCAATAGCTTCAAG | 2207 | | | | | | |
| DB | 2067 | AAGCTCTTTAAGCTGATGAGCACTTCTGCGCGCTTCTGAATGATACTCGAACTTTGAT | 2126 | | | | | | |
| QY | 2208 | AGGGAATTTAAGGAAGCAAAATTTAAACGCGGTAGCATTTGATTTAGTAAACGAGAAAGT | 2267 | | | | | | |
| DB | 2127 | AGAGTCCAGGAGGGAATTTAAATGCTTCTCTATACATGATTTAGTCCGCTGCT | 2186 | | | | | | |


```
Db 1664 GGCAGGACCCCTTATCGATCCTGAACTTCATGCTCGCATAGCATGGGCACAAAATGG 1723
QY 1736 CATATTAACATAAGTAGTGTGATGACATCTTTTGTGATTCGGTGTGACATGCAATGAC 1795
Db 1724 TGTGCTCAGACCGTGGTGTGATGATCTTATGATGGTGGAGATCTGAAAGAGAAATGGA 1783
QY 1796 CAACCTGATCAATGTGTGAAAATGGAATGTAGATGTCGACAAAGGATTTGTTTCAGA 1855
Db 1784 TAACTCTTATAGAAATGTGGTGGAAAAGTGGATCCTGATGGGGAAGTGGTTACTGTTCCA 1843
QY 1856 GCAATGCTCGGATTTTATTTTATAGCAATTAAGATGCAATCTGTGTGATTTGGAGATGAAC 1915
Db 1844 GACGTTGAGATGTATTTCTTGCATGTCACACACAGTTTGTGAAATAGGAAGAGAGC 1903
QY 1916 TTTTAAATGGCAAGCCGGATGTAACTAGCCATGTTATTTCAAACTTTGGTTGGAACTAAT 1975
Db 1904 TTTAGTATGGCAAGGACCGAGTGTATGAGGAATGTTATCGATGGTGGTGGCTCTGCT 1963
QY 1976 GAATAGTATGTTGAGAGAGCTATATGACAAAGATGCTTATGTGCCAACATTAATGA 2035
Db 1964 GAAGGTGATGAGAAAGGAAGCTGAATGGTGCACAAATTAAGGTAGTGGCAATCAATGGTGA 2023
QY 2036 ATATAGTGAAGAACGCTTACGTGTCAATTTGCAATTTAGCCCGATTTGTCAAGCCGGCTATTA 2095
Db 2024 ATATATGGAACAAGCCCATGTATCAATTCGGTTGGGACCTATATAATCCTTCCAATGCTCT 2083
QY 2096 CTTTGTGGGGCCCAATATATACAGAGAGATGTTGAAAGCTCTGAATATCAATATCTAAT 2155
Db 2084 CTTTGTGGACCTAAACTCTCAGAGGAAATGATTTGGAAGCTGTGAATACCAAGATTATA 2143
QY 2156 TAAGCTAATGACACGAGGCTCGACTTCTTAAACGATATCCATGCTTCAAGAGGGAAT 2215
Db 2144 TAAGCTGATGAGACACTGTGTCGCTTAAAGATGATGATGATGATGATGATGATGATGATG 2203
QY 2216 TAAGGAAGGCAATTAAGACCGGTAGCATTTGCAATTTGAGTAAACGGAAGTGGGAAAGT 2275
Db 2204 CAAAGAGGGAAGCTGAATATCTGCTCTGTGATGATGATGATGATGATGATGATGATG 2263
QY 2276 GGNAGAAGAGTGTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2335
Db 2264 CAAAGAGGAGGCAATTAAGACCAATTAAGAGGGAATTTGAGAGGCGGATTAAGAGAGCTG 2323
QY 2336 GAAATTAATTTTGAAGAAATGGTAGCATTTGCTTCTAGAGCTTTGTAAGATGCAATTTG 2395
Db 2324 GGGTTAGTTTTCAGGAGACACTACAA---TTCCAAGAGCTTTGAAGGATTTGTTCTG 2380
QY 2396 GAACATGTGTCAAGTGTGAAATTTTTCAGCAAAAGATGACGGGTTTACTGGAAACAC 2455
Db 2381 GAAATTAATGATCCATTTGTAATCTATTTTACATGGAAGATGATGGGTACACTTCAAATAG 2440
QY 2456 GATCTGTGATGATGTAAGAGACATCTTACACACCGTGGTGGCTTG 2502
Db 2441 GTTGATGAACACTGTAAAGCCATGTTTGAACCAACCCATGGATCTGG 2487
```

```
RESULT 6
US-10-041-018-381
; Sequence 381, Application US/10041018
; Publication No. US2004007232A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080051/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 2658
```

```
; TYPE: DNA
; ORGANISM: Cucurbita maxima
; US-10-041-018-381

Query Match      28.0%; Score 782.2; DB 16; Length 2658;
Best Local Similarity 61.3%; Pred. No. 1.7e-173; Indels 54; Gaps 6;
Matches 1390; Conservative 0; Mismatches 823;

QY 275 AACTATATCCCACTAATCTGATCATTTGATACAAACGATCCCAAAACAGTTTAA 334
Db 236 AACTATAAATGAGCTTTGCACTTTGAAGAAACAAAGAAATTAAGAAATTTGCGA 295
QY 335 AATGTAGAAAATTTCTGTTTCTTCAATGACACAGATGGGTAGCCATGGCTCTCTCC 394
Db 296 CAAGGTTGAACCTTTTCAGTTTCTGCAATGATGATGCGGTGCGCAATGGTCTCTCC 355
QY 395 AAATCTACCCCAATCGCTTGTTCCTTGAGTGTCTCAATTTGGTTAAATTAATACAGCT 454
Db 356 AAATCTCTCTCAACCAACCTCTTTTCCCGAGTGTATAAACTGGTATTTAGATAGTCA 415
QY 455 TAATGATGTTTCATGGGGTCTTGTGTTAAATCACACTCATAATCATATCAACCGGTGCTTAA 514
Db 416 TGTGATGCTCATGGGGCTTACTCCACAAC-----GATCAGTTGCTGATGA 463
QY 515 AGATTTCTTATCTTCAACATTTAGCATGATTTGTCATTTAAAGATGGAATTTGGGA 574
Db 464 GGCCAACTCTTATCTACATTTAGCATGTTCTTACTCTTAAGCGGTGGAATATTTGGGA 523
QY 575 AGATCAATTAATTAAGGTCTAAGTTTATTTAGTCAATCTTGGTTCAGTCTACTGAAA 634
Db 524 TGATCATATAGCAAGGCGCTTGAATTTATTAAGTCTAATATAGCTTCACTACTGATGA 583
QY 635 AAGTCAACCATCTCCCAATTTGGTTTTCAGATCATATTTCTCTGGTTTCTTGGATGCGAA 694
Db 584 GACCAACGTTCTCCGTTGGGATTTGACATTAATTTCCCTGGCATGATTTGATGCTAA 643
QY 695 AAATTTGGACATAAACCCTCTTCAAAACAAACAGATTTTGTGATGCTACATAGAG 754
Db 644 AGACTTTGAATTTGAATCTACCTTGGCACCGACGAAACGTTGGTTCGTAAGAA 703
QY 755 GGAATTTGGAGCAAAA-----AGATGCCATTCAAATGAGATGGATGATACATTTGGCGTA 808
Db 704 AGATTTGGAGCTGAGAGCTGAGAGCAACTCTGAGGTGGAAAGCCATTTTAGGGTA 763
QY 809 TATCTCTGAAGGACTCGGTAATTTATGATTTGGAATATGGTGAAGAAATATCAGATGA 868
Db 764 TGTTCAGAGGAATTTGGAAGTTACAGGACTGGGATGTTGTCATGCAATATCAAGGAA 823
QY 869 AAATGTTCTGTTTCAACTCAGCATCAGCAAGCTGCTGCTTTTCAATTAATCATCAAAA 928
Db 824 GAATGGATCACGTTTAAATTTCTCCATCCACTACGGCAGCGGCTTTTATGCAATAGAAATGA 883
QY 929 TCCGTGTTCTTAATTTTAAATTTCACTTTTGGACAAGTTTGGTAATGCAGTCCCAAC 988
Db 884 TGATGGCTGTTTGAATTAATCTCGCTACCTCTTACAAAAGTTTGTAGTGGCTCAGTCCAC 943
QY 989 AGTTTATCTCATGATTTATTTATCCGACTTTTCTATGTTTGAACACAAATTTGAAGATPAG 1048
Db 944 AATATATCTCTTGATATATATCTCGATTTACACATGTTTGTATAGCTTCAAAAATTCGG 1003
QY 1049 AATTCACACATTTGAGTGAATTAATAATTTTATAGATGAACACATACAGATGTTG 1108
Db 1004 AATTTGCTCGGCTTTCAAGAGAGGATTTAGAACGCTATTTAGTGAACCTTTACAGGTTG 1063
QY 1109 GGTGGAACGAGATGAGCAATATTTTATGATGATGTTTGAACATGTTGCTTTTACGCTTTT 1168
Db 1064 GATGGAAGGAGAGAAATATATTTCTTAGATGTTCTCACTTTGTGCAATGGCTTTTGAAT 1123
QY 1169 ATTAAGGATCAATGGGTATGAAGTTTCCAGATGCCATTTGGCTGGAATTTACTAA----- 1222
Db 1124 GTTACGGTTGAAGGATATGATGTTTCTTACAGACCAAGTTTCACTCAATTTTCAAGATAT 1183
QY 1223 -----TGAAATTAGCTTTGAAGAGCATATATGACGCTCTTGAACATATCA 1267
```

Db 1184 CTTTCCCAATTCCTTGGAGGATATTTAAAGACATTCGGTGGCTCGCTGGAGTTATATAA 1243
QY 1268 TCGCTCACATATATATACCAAGAGGATTTACTTCT---GGAAACAATCTTGAAGTC 1324
Db 1244 GGCCTCTCAGATATACACGCCCGGATGAATCTGTTCTGGAAAAATATAAACTCTTGGAC 1303
QY 1325 AGCTGATTTCTTCAAGAGATATATATCACTGATTCAAACAGGCTTTCATAA----- 1376
Db 1304 TAGTCGTTTCTTGAAGCATGATTAATCTAGTGTGATTCAGTTGGTCTGATAGACCGATAG 1363
QY 1377 -TTAATTCACAAGAGGTGGAAAAATGCTCTTAAGTTCCCTATCAATACCGGTTTGAACG 1435
Db 1364 TGTGTTTAAACAAGAGGCTTGAATGCTCTTGAAGTTCCCTATTAATGCAACTCTAGAAGC 1423
QY 1436 CATAAACACTAGACGAATATATACAGCTTTACAACTGTAGCAATACAAAGATTTCTGAAAC 1495
Db 1424 CCTAATAGTAGAGGGCAATGGAAGTTACAGTGGAGACATGTGAGGATTTCAAAATC 1483
QY 1496 TACATATCACTCATCAATATATAGTAACACTGATTACCTTAAGTTGGCTGTGTAAGATTT 1555
Db 1484 GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGGAACCTGCTGTAGAGATTT 1543
QY 1556 CTACACCTGCCAATCTATTTATCTGTAAGAAATTAAGTCTTGAAGGTGGTGGTGA 1615
Db 1544 CAATACCTGCAACGCATTTCTATTAAGAACTGGAAGAGCTTCAAGAGATGGTGGTGA 1603
QY 1616 GAATAAGTTGGACCAAGCTCAAGTTTCTAGGCAAAAGACCGCTACTGTTTCTGTT 1675
Db 1604 AACAATTTGACAGATTTGAATTTTTCAGCTGCACCTGAGGTACTGCTATTTTGTGTC 1663
QY 1676 TGTGCAACACTTTGCTCCGATATATCAGATCGCGCTATTTTCAAGGGCCAAATAAG 1735
Db 1664 GGCAGCGACCTTACTGATCTGAACTTCATGATCTCGCATAGCATGGGACAAATGG 1723
QY 1736 CATATTAACACAGTAGTTGATGATTTTGTGATATCGTGGGTACAAATCGATGAATGAC 1795
Db 1724 TGTGCTACGACCGTGGTGTGATGTTTCTATGATGCTGGAGGATCTGAAGAGGAATGGA 1783
QY 1796 CAACCTGATTCAGATGTTGAAAAATGAAGTATGATGTCGACAGGATTTGTTTGA 1855
Db 1784 TAACTTATAGAAATGTTGGAAGTGGGATCTGATGGGGAAGTGGGTACTGTTCCAA 1843
QY 1856 GCATGTTGGATTTTATTTTAGCATTTAAAGATGCAATCTGTTGGATTGGAGATGAAGC 1915
Db 1844 GGAAGTTGAGATTTGTTTCTGCACTGCACAGCACAGTTTGTGAATAGGAAGAGGC 1903
QY 1916 TTTTAAATGGCAAGCGCGATGTAACCTAGCCATGTTAATCAAACTTGGTTGGAATAAT 1975
Db 1904 TTTAGTATGGCAAGGACGACGTGTTATGAGGAATGTTATCGATGTTGGTTGGCTCTGCT 1963
QY 1976 GAATAGTATGTTGAGAGAAGCTATATGACAGAGATGCTTATGTCGCAATTAATGA 2035
Db 1964 GAGGTGATGAGAAGAGAGCTGAATGTGCAAAATAGGTAGTGGCATCAATGGGTGA 2023
QY 2036 ATATATGAAACCCCTTACGTTGATTTGCAATAGCCCGGATTTGTCAGCCGGCTATTTA 2095
Db 2024 ATATATGAAACAAGCCCATGATCATTCGCGTTGGACCTATATCTTCAATGCTCTT 2083
QY 2096 CTTTGTGGGCCCCAAATATCAGAGGAGATTTGTTGAAGCTCTGAAATATCAATATCTATT 2155
Db 2084 CTTTGTGGAACCTTAACTCTCAGAGGAATGATTTGGAAGCTGTGAATACCAAGATTTATA 2143
QY 2156 TAACTAATGAGCAGCGAGGTGCACTTCTAAGCATATCATAGCTTCAAGAGGAAT 2215
Db 2144 TAAGCTGATGAGCACTGCTGTGCTGCCCTTAAAGATGATATTGCAFTTACGATAGAGAATG 2203
QY 2216 TAAGGAAGGCAAAATTAACCGGTAGCATTTGCAATTCAGTAACGAGAAAGTGGGAAGT 2275
Db 2204 CAAAGAGGAAAGCTGATATTTCTGCTCTGATGATGATGATGCGGTGATGTCAC 2263
QY 2276 GGAAGAAGAGGTTGTGAGAGATGATGATGATGATTTAAACAAGAGGAAAGATTAAT 2335

RESULT 7

US-10-041-018-377
; Sequence 377, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 377
; LENGTH: 2506
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-041-018-377

Query Match 26.2%; Score 732.4; DB 16; Length 2506;
Best Local Similarity 59.5%; Pred. No. 9.2e-162;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

QY 270 GGACAAACTAATCCACATACTCTGATCATTTGATACAAAGAACGGATCCAAACACAG 329
Db 143 GTACAGACAAGAGCTAACAATGTGAGCTTTGAGCAACAAGAGAGAGATTAGGAAGTG 202
QY 330 TTTAAAATGTAGAAATTTCTGTTTCTCATATGACACAGCATGGTAGCCATCGTCCCT 389
Db 203 TTGAGAAAGTGGAGCTTTCTGTTTCCGCTTACGATAGTCTAGTTGGTAGCAATGTTCCA 262
QY 390 TCTCCAAACTCACCCAAATCGCTTGTTCCTCGAGTGTCTCAATTTGGTTAATTAAT 449
Db 263 TCACCGAGCTCCCAAAATGCTTCCACTTTTCCACAGTGTGTGAATGGTTATTGGATAAT 322
QY 450 CAGCTTAATGATGTTTCATGGGTCTTGTGTTAATCACACTCATATCATATACCCGTTG 509
Db 323 CAACATGAAGATGATCTTGGGACCTTGATACC-----ATGACCATCATCTCTT 373
QY 510 CTTAAGACTTCTATCTTCAACATGATGATGATTTGCAATTAAGAGATTAAGAGATGTT 569
Db 374 AAGAAGGATGTTTATCATCTACACTGGCTAGTATCTCGGTTAAAGAGATGGGAAT 433
QY 570 GGGGAAGATCAATAAATAAGGCTTAAGTTTATTTGAGTCAATCTTCTGCTAGCTACT 629
Db 434 GGTGAAGAACAATAAACAAGGCTCTCCAGTTATTGAGCTGAATTTGATAGTCACT 493
QY 630 GAAAAAGTCAACCATCTCCCATTTGTCATCATATTTCTGCTGTTTCTGTTGAGTAT 689
Db 494 GATGAACCATACAGAAACCAACAGGTTTGAATTAATTTCTGGAGATTAATAAT 553
QY 690 GCGAAAACTTGGACATPAAACCTCTCTTTCAAAACAAACAGATTTTACTTTGATGTACAT 749
Db 554 GCTAGAGATTTGAATCTGACGATTCATTCGCTCGAGAGTGGTGGATGACATGATACGA 613
QY 750 AAGAGGGAATTGGAGCAAAAAG-----ATGCCATTCAAATGAGATGGATGATAC 800

Db 614 AAAGAGATCGGATCTTAATGTCATAGTGAAAGATTTTCAAGGAGAGACATAT 673
Qy 801 TTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATAGATGGAATATGGTGAAGAAATAT 860
Db 674 CTGGCCTATGTTTATAGAGGGAACAAGAACTTAAAGATGGGATTTGATAGTCAATAT 733
Qy 861 CAGATGAAATATGGTCTGTTTCACTCACTCAGCAACAGCTGCTCTTCAATTAAT 920
Db 734 CAAAGGAAATATGGGTCACTGTTGATTTCCAGCCACACAGCAGCTGCTTTTACTG 793
Qy 921 CATCAAAATCTGGTGTCTTAATTAATTTAAATTTTCACTTTTGGCAAGATTTGGTAATGCA 980
Db 794 TTTGGGAATGATGGTGTCTCGTATCTCTGTTCTCTCCTTTCAGAAATTCGAGGCTGCA 853
Qy 981 GTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGGTTCACAAATGAA 1040
Db 854 GTTCTTTCAGTTTATCTCAATTCATTAATATGACGCTTTAGTATTAATTTGCTCACTCTTGA 913
Qy 1041 AGATTAGGAATTTCAACCAATTTTCAGAGTGGAAATTTAAATATGTTTATGATGAACATAC 1100
Db 914 AGCTTAGGAATTTGATAGATTTTCAAAACCGAAATTCAAAGCATATTTGGATGAACCTAT 973
Qy 1101 AGATTGTTGGTGGAAACGAGATGAGCAAAATATTATGATGATGTTGTAACATGCTTTAGCC 1160
Db 974 AGATATTGGCTTCGTGGGATGAGAAATATGTTTGGACTTGGCCACTTGTGCTTTGGCT 1033
Qy 1161 TTTCGGTATTAAGGATCAATGGGTATGAGTTTCCCGAGATCCATTTGGCTGAATTTACT 1220
Db 1034 TTCCGATTTATGCTTCTCATGGCTATGATGTTCTTTAGCATCCGCTAAACCAATTTGCA 1093
Qy 1221 AATGAATTAGCTTTGAAAGCAATATGAGCTTTTGAACATATCATCGCTCACATATA 1280
Db 1094 GAAGAATCTGGTTTCTGATATCTTGGAAAGGATATGTTAAGATAGCTTTCTGTGTTA 1153
Qy 1281 TTATACCAAGAGAT-----TTATCTTCTGGAAACAAATCTTGAAGTCAAGCTGATTTCC 1335
Db 1154 GAATTTATTAAGCTCTCBAAGTTATCCACATGAATCAGCTTTGAAGAAGCAGTGTGT 1213
Qy 1336 TCAAGAGATATATCCACTGATCAA-----ACAGGCTTTCT 1373
Db 1214 TGGACTAAACATATCTGGAGATGGAATTTGCCAGCTGGGTTAAGACCTCTGTTGAGAT 1273
Qy 1374 AAATTAATTCACAAAGGTTGAAATGCTTCTTAAGTTCCCTATCAATACCGGTTTAGAA 1433
Db 1274 AAATACCTCAAGAAAGGTCGAGATGCTTCTTCTTTCCCTCTTCCCTATGCAACCTAGAA 1333
Qy 1434 CGCATAAACACTAGACGAATATAC---AGCTTTACATGTAGACAAATACAAATTCGT 1490
Db 1334 AGATCAGATCACAGGAGAAATACTCAATGTTCTGCTGTGGAAACACACGAGTTACA 1393
Qy 1491 AAACTACATATCACTCATCAATATATAGTAACTGATTAAGTTAGGTTGGCTGTTGA 1550
Db 1394 AAAACCTCATATCGTTTGCAATATTTGCACCTCTGATATCTGGAAGTTAGCTGGAT 1453
Qy 1551 GATTCTTACACCTGCCAATCTATTATTCGTAAGAAATTTAAAGCTCTTGAAGTGGGTG 1610
Db 1454 GACTTCAATTTCTGCCAGTCCATACACCGTGAAGAAATGAAAGTCTTGTATAGTGGATT 1513
Qy 1611 GTAGAGAATAAGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTGTTATTC 1670
Db 1514 GTGGAGAATAGATTGCAAGGAACGAAATTTGCCAGACAGAAGCTGCTTACTGTTATTC 1573
Qy 1671 TCTGTTGCTCAACACTTTGCTCTCCGAAATTCAGATCGGCTATTTTCATGGGCCAAA 1730
Db 1574 TCTGGGGCTGCAACTTTATTTTCTCCGAACACTATCTGATGCTGATATCTGTTGGCCAAA 1633
Qy 1731 AATGGCATATTAACACAGTAGTTGATGACTTTTTTTCATATCGGTGTGATCAATGATGAA 1790
Db 1634 GGTGGAGTACTTACAGGGTTGTAGACGACTTCTTTGATGTTGGAGGTTCCAAAGAGAA 1693
Qy 1791 TTGACCAACTGATTTCAATGTTTGAAGAAATGGAATGTAGATGTCGAAGGATTTGTT 1850

Db 1694 CTGGAACCTCATACACTTGGTGAAGGATGGGATTTGAACGGTGTTCCTGAGTACAGC 1753
Qy 1851 TCAGAGCATGTTTCGATTTTATTTTATAGCATTTAAAGATGCAATCTGTTGATTTGGAGAT 1910
Db 1754 TCAGAACATGTTGATCATATTTCTCAGTTCTAAGGGACACCAATTTCTCGAAACAGAGAC 1813
Qy 1911 GAAGCTTTTAAATGGCAAGCGCGCATGTAACCTAGCTGTTATTTCAAACTTGGTTGGAA 1970
Db 1814 AAAGCAATTCACCTATCAAGGACGCAATGTGACACACCAATTTGTGAAATTTGGTTGGAT 1873
Qy 1971 CTAATGAATAGTATTTGAGAGAGCTATATGGACAAGAGATGCTTATGTGCAACATTTA 2030
Db 1874 CTGCTCAAGTCTATTTGAGAGAGCCGAGTGGTCCAGTGACAAGTCAACACCAAGCTTG 1933
Qy 2031 AATGAATATATGAAAAACGCTTACGTGTCATTTGCAATTAGGCCCGGATTTGCAAGCGGT 2090
Db 1934 GAGGATTTACATGAAAAATTCGTACATATCAATTTGCTATAGGACCAATTTGCTCCAGCT 1993
Qy 2091 ATTTACTTTTGGGGCCCAAAATTATCAGAGAGATTTTGAAGCTCTGATATCATATAT 2150
Db 1994 ACCTATCTGATCGGACCTCCACTTCCAGAGAAAGACAGTCCGATAGCCCAATATAATCAG 2053
Qy 2151 CTAATTAAGCTTAATGAGCAGCAGGTCGACTTCTTAAACGATATCCATAGCTTTCAAGAG 2210
Db 2054 CTCTCAAGCTCGTGAACACTATGGTCTCTTTTAAATGACATCAAGGTTTAAAGAA 2113
Qy 2211 GAATTTAAGAAAGCAAAATTTAAACGGGTAGCATTTGAGTAAACGAGAAAGTGGG 2270
Db 2114 GAAAGCGGAGGAGGAGCTGATGCGGTTTCATTTGCAATGAAACACGAGAGAGACAAT 2173
Qy 2271 AAAGTGGAAAGAGAGGTTGTGAGAGATGATGATGATGATTAATAAAACAGAGGAGAA 2330
Db 2174 CGCAGCAAGAAAGTGCATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2233
Qy 2331 TTAATGAATTAATTTTCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2390
Db 2234 TTGCAATAGCTAGTTTGGAGAGAAAGGATGTTGTTCCAGGGAATGCAAGAGAGCG 2293
Qy 2391 TTTTGAACATGTCACGTTGCACTTTTAAATTTTAAACGCAACGATGACGGGTTTACTGGA 2450
Db 2294 TTCTTGAATAGCAAAAGTGTGAACTTATTTTACAGGAAGGACGATGATTCACATCA 2353
Qy 2451 AACACGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2510
Db 2354 AATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2413

RESULT 8

US-10-425-114-3441
; Sequence 3441, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3441
; LENGTH: 2554
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700265067_FLI
US-10-425-114-3441

Query Match 20.1%; Score 561.2; DB 16; Length 2554;

| Best Local Similarity 56.1%; Pred. No. 2.1e-121; Matches 1210; Conservative 0; Mismatches 893; Indels 53; Gaps 6; | | | |
|--|--|---|------|
| QY | 324 | AAACAGTTTAAATAATGATAGAAATTTCTGTTCTTCAATACACAGCAGTGGTACCGATG | 383 |
| Db | 160 | AATCAGCTCCGGAAGCCCGAGTTGCCACTTCTTATACACACAGCGTGGTCTCTATG | 219 |
| QY | 384 | GTCCCTTCTCCAAACTCACCAATCGCCTGTGTTCCCTGAGTGTCTCAATTTGGTTAATT | 443 |
| Db | 220 | GTGCCAGTGGGGCTCTCATCAGACTCCCGCTTCCCACTGTGTTGAGTGGATATTG | 279 |
| QY | 444 | AATATCAGCTTAATGATGGTTTCATGGGTCT-TGTATATCACTCATTAATCAATCA | 502 |
| Db | 280 | CAGAAATCAGCAGGATGATGGATCTTGGGTGTCCAGCAATCTGACTCAT----- | 328 |
| QY | 503 | CCCGTTGCTTAAAGATTCTCTATCTTCAACATTAGCATGTATTGTGCAATTAAGAATG | 562 |
| Db | 329 | --CGTACAGAGGATGTTCTCTATCCACGTGGCATGTGTCTTGGGTGAAGAGATG | 386 |
| QY | 563 | GAATGTGGGGAAGATCAAAATAAATAAAGGCTAAAGTTTATTAGTCAAAATCTTGTCT | 622 |
| Db | 387 | GAATGTGGCAGAGAGACATTTGGAGAGACTGCATTTTCATCGGAGGAATTTCTGT | 446 |
| QY | 623 | AGTACTGAAATAAGTCAACCATCTCCCATTTGCAATTTTTCATCAATTTCTGTGTTGCT | 682 |
| Db | 447 | TGCTATGGACCAAGTTCACCTCTCTATAGTTTCAACATCACCTTTCTGTGTTGCT | 506 |
| QY | 683 | TGAGTATCGGAAATACTTGGACATAAACCTCTTTCAAAACAAACAGATTTTAGTTGAT | 742 |
| Db | 507 | TAACTCGGCATTGATATGGGTGTAGAAATTTCTGTAAAGACAAACTGATGTCTGTGGCAT | 566 |
| QY | 743 | GCTACATAAGAGGAATTTGGAGCAAAAAGATGCCA-----TTCAAATGAGATGA | 793 |
| Db | 567 | TCCTCACCGCGGAGATGGAATTCGAAAGGCTGCTGTGATGTCTTTTGGAGAAA | 626 |
| QY | 794 | TGATATCTGGGTATATCTCTGAGAGACTCGTAAATTTATGATTCGAAATATGGTGAA | 853 |
| Db | 627 | AGCATATATGGCTTTATCCAGAGAGATTCGGAATATGCTGGACTGGGATCAAGTTAT | 686 |
| QY | 854 | GAATATCAGATGAATAAGTCTGTGTTTCACTCACCATCAGCAACAGCTGCTGCTTT | 913 |
| Db | 687 | GAAGTTTCAGAGGAGAAATGATCAATGTTGAGCAGTCTTCCCAACTGCTGTGTGAT | 746 |
| QY | 914 | CATTAAATCATCAAAATCTGTTGTCTTAATTTAATTTCACTTTTGGACAGTTTGG | 973 |
| Db | 747 | AATCCACAAATACACGACCAGCCCTTCATACCTAAATTTGTTGTGATGATTTGG | 806 |
| QY | 974 | TAATGCAATCCCAACAGTTTATCTCATGATTTTATTTATCCGACTTTCTATGTTGACAC | 1033 |
| Db | 807 | CAGTGCAGTACAGCAATGTATCTTCAATGGTACATTTGTGAGCTTTCAATGGTGGACG | 866 |
| QY | 1034 | AATGAAAGATTAGGAATTTTACACCACTTCAGAGTGGAAATTAATAATGTTTAGATGA | 1093 |
| Db | 867 | GCTTGAATAAATGGAATTTCTCAGCGCTTGTGAGTGAATAAGAAAGCAATCTGGACAT | 926 |
| QY | 1094 | AACATACAGATTTGGGTGGAAACAGATAGCAATAATTTCAATGATGTTGTAACATGTC | 1153 |
| Db | 927 | GGCATACAATTTGCTGTTTACAGAAAGATGAGGAATCATGATGACATAGCAACATGTGC | 986 |
| QY | 1154 | TTTAGCCCTTTGCTTATTAAGATCAATGGGTATGAAAGTTTCCCAAGATCCATTTGGCTGA | 1213 |
| Db | 987 | AATGGCAATTTGCGCTTTTGGAGTGAATGGTTTACATGTTTCTCTCAGATGATGTCTCA | 1046 |
| QY | 1214 | AATTAATAATG-----AATAGCTTTGAAAGCAATATGC | 1249 |
| Db | 1047 | CGTTGCTGGAGCTTCCACTTTCCATGATTCATCAAGAGGATTTTAAATGATACAAAATC | 1106 |
| QY | 1250 | AGCTCTTGAACATATCATCGTGCATATATTAATACCAAGAGATTTTATCTTCTGGAAA | 1309 |
| Db | 1107 | CCTACTGGAAATGTAAGAGCTTCAAGTCAACCTTATCAGAAACGATCTGATCTTAGA | 1166 |
| QY | 1310 | ACAAATCTTGAAGTCACTGATTTCTTCAAGAGATAATATCCACTGATTTCAACAGGCT | 1369 |
| 1167 | TCGCATAGTTTCTCGTCTGGCAACTTATTGAAGGATTAAGATGTGCTGTAGTAGGTGCA | 1226 | |
| 1370 | TTCTAAATTAATTCACAAAGAGTGGAAATCTCTTAAGTTCCCTATCAATACCGGTTT | 1429 | |
| 1227 | AAAAGACTCGAATTTTGGAGAGATGGAGTATGCTGTAAATTTTCCCTTGTATCCACACT | 1286 | |
| 1430 | AGAAGCAGATTAACACTAGACGAAATATACAGCTTTACAAATGTAGACAATAAAGATTTCT | 1489 | |
| 1287 | GGAGCGCTAGAAACAAGAGAGAAACATCGAAATTTGTATGCTTGGGGTCTCTCATGTCT | 1346 | |
| 1490 | GAATACTACATATCACTCATCAAAATATTAGTAACACTGATTACCTAAGCTTGGCTGTGTA | 1549 | |
| 1347 | ---AACACAAATCTCTCATCTTTTCGTATCAATCAAGAAATTCCTAGCTTGGCAGTGA | 1403 | |
| 1550 | AGATTTCTACACTCCCAATCTATTTATTCGTGAAGAATTAAGAAGTCTTTGAAGGTGGT | 1609 | |
| 1404 | AGATTTCAATTTCTCTCAACGTTGTTTACCGGATGAACTTCGGCATCTTTGATGTTGGT | 1463 | |
| 1610 | GGTAGAATTAAGTTGGACCAAGCTCAAGTTTGTCTAGGCAAAAGACCGCTACTGTATTT | 1669 | |
| 1464 | GAAGAGAACCAAGCTGGACCACTACAAATTTGCTGGCAGAAACCTGACATATTGCTATCT | 1523 | |
| 1670 | CTCTGTTCTGCAACACTTTCTGCTCCGAAATTAATCAGATGCGGTATTTTCATGCGCCAA | 1729 | |
| 1524 | GTCTGCTGCTACCGTATTTTCTTGAATTTGCTGACCTCGCATTTCTATGGGCCAA | 1583 | |
| 1730 | AAATGGCAATTAATCACTACAGTAGTTGATGACTTTTTCATATCGGTGGTACAAATCGATGA | 1789 | |
| 1584 | AAATGGTGTCTCAACTGTGTTGATGACTTCTTCGATGTGTTGGATCAAAAGAGA | 1643 | |
| 1790 | ATTGACCAACCTGATTCATGTTGTTGAAATTAAGATGATGTGCGACAAGATGTTGT | 1849 | |
| 1644 | ATTGAAACCTGATAGCACTAGTTGAGAAATGCGCATGCGCACCATGCAATGCTGATCTTA | 1703 | |
| 1850 | TTGAGAGCATCTTCGGATTTTATTTTTAGCATTAAGATGCAATCTGTTGATTTGGAGA | 1909 | |
| 1704 | TTGGAACAGTGAATAATAGTATTTTCTGCTATTTTATCAACAGTGAACCATCTTGGAGC | 1763 | |
| 1910 | TGAAGCTTTTAAATGGCAAGCGCGATGTAACCTGATGCAATCTTATTTCAAACTTGGTGA | 1969 | |
| 1764 | AATGGCTTCTGACGACAAGCGCGTATCTTACAAACCACTAGTAGAATAATGGCTGGA | 1823 | |
| 1970 | ACTAATGAATAGTATTTGAGAGAGCTATATGACAGAGATGCTTATGTGCCAACAT | 2029 | |
| 1824 | TTGTTAAGATCTATGATGTTGCGAGGAGATGCGAGAGATGCCAATATGTACCAACAGT | 1883 | |
| 2030 | AAATGAATATATGAAAAACGCTTACGTGTCATTTGCAATTAGCCCGGATTTGCAAGCCGC | 2089 | |
| 1884 | TGAAGATACATGACAAATGCTGTTGCTCATTTGCACTGGGCCCAATTTGCTCCACG | 1943 | |
| 2090 | TATTTACTTTGTTGGGCCCAATTTATCAGAGGAGATTTGTTGAAAGCTCTGTAATATCATAA | 2149 | |
| 1944 | ATTGTATTTGTTAGGCAAGAGCTATTAGAGCATGCTGTCAAAGATGAAGTAGACATAA | 2003 | |
| 2150 | TCATTTAAAGTAAATGACGCGAGGCTGCACTTCTTAAAGATATCCATAGCTTCAAGAG | 2209 | |
| 2004 | ATTTATTTAGCTAGTAGAGCTTTCGGGAGGCTCTCAATGACTACCAAAAGTTTAGAGAG | 2063 | |
| 2210 | GGATTTTAAAGCAAGCAAAATTAACGCGGTAGCAATTCATTTGATTAACGAGAGAAATGG | 2269 | |
| 2064 | GGAGGCAACGAGGAGAGCTGAAATGTTTCTTACTTGTGCTCCACAGTGTGTTTC | 2123 | |
| 2270 | GAAAGTGGAAAGAGGTTGTGGAGGAGATGATGATGATGATTAATAAACAAGAGGAAAGA | 2329 | |
| 2124 | TATGTCCTCAGAGCCGCTAAAAAGGCAATGACAGAGTCCATAGACGCTGTCTAGGAGAGA | 2183 | |
| 2330 | ATTAATGAAATTAATTTTGAAGAAATGATAGCATTTGTTCTCTAGAGCTTGTAAAGATGC | 2389 | |
| 2184 | CTTGTAAGATTTGGTTCT---CAGGAAAGAAAGTGTGTTTCTTAGGCCATGCAAGAGCT | 2240 | |
| 2390 | ATTTTGGAAACATGTGTCAGCTGTGATTTTATTTTATAGCAACGATGACCGGTTTA | 2445 | |
| 2241 | CTTCTGGAAGATGTGAAGTACTTCACTGTTTTTCTCAGATGATGGAATTTA | 2296 | |

Db 719 GCTCTGATGGA---CGCAAGAGTACCGATATCTTGAAGAGGTAGATATGCCCTTAAGT 775
QY 1411 TCCCTATCAATACCGGTTTAGAAACGATATAACACCTAGACGAATATACAGCTTTTCAATG 1470
Db 776 TTCCCATTTAGCCCTAGTAGATCTCTCAACAACAGAGAAGTATGGAGCAATTTGGATG 835
QY 1471 TAGACAATACAGAATTTCTGAATACTACATATCACTCATCAATATTTAGTAACATGATT 1530
Db 836 CTAGGGGTTCTCAGACGGCTAAAGACAAATATCTCCATGTCTCAATGTAAGTCA---AGACA 892
QY 1531 ACCTAAGGTTGGCTGTTGAAGATTTCTACACCTGCCAATCTATTTATCTGTGAAGATTA 1590
Db 893 TCCTTGGCTGGCTGTGAAGATTTCTGTTTTCTCAATCTATTTTACCGAGGAGAACTAC 952
QY 1591 AAGGTTCTTGAAGGTTGGGTGAGAGATAAGTTGGACCAAGCTCAAGTTTGTAGGCAAA 1650
Db 953 AGAATCATTTAGTTGGGAGAGAGAGATAGGATGGACCAAGCTCAAAATTTGTGCGCAAA 1012
QY 1651 AGACCGCTTACTGTTATTTCTCTGTTGCTGCAACTTTCTGCTCCCGAATTTATCAGATG 1710
Db 1013 GGTGCGCATATTGCTATCTCGCTGCTTACCACCATATCCCTCATGAATTTGCTGTGATG 1072
QY 1711 CGCGTATTTTCATGGCCCAAAATGGCATATTTACTACAGTAGTTGATGACTTTTGTGATA 1770
Db 1073 CTCGGTTGCTATGTCACAAAGCATCATGCTCAGGTTGATGTTGATGACTTTCTTCGATG 1132
QY 1771 TCGGTGTACAAATCGATGAATTCACCAACCTCATTCATTCATTTGAAAAATGGAATGTAG 1830
Db 1133 TTGTGTGATCAAAAGAGAGAAATCTCATCGAATTAGTTGAGAACTGGATGAGC 1192
QY 1831 ATGTGCAAGGATTTGTTTTCAGAGCATGTTCCGGAATTTTATTTTATTTAGCATTAAGAATG 1890
Db 1193 ACCACAAAGTTGAGTTCTCTGTCGGAGAAAGTAAATAGTTTCTATGCTGTCTATAATA 1252
QY 1891 CAACTCTGTTGGATGAGATGAAGCTTTTAAATGCGCAAGCGCGCATGTAACTAGCCATG 1950
Db 1253 CAGTGAACGAGCTTGGATCTATGCGCTTCTGAGTACAGAGCCGATGTCACAAACACC 1312
QY 1951 TTATTCAAACTTTGGTTGGAATTAATGAATAGTATGTTGAGAGAAGCTATATGACCAAGAG 2010
Db 1313 TCGCTGAACATGGCTTAAAGTAATGCTGTGATGCTGACGAGGAGAGAGCTGCGCAAGGA 1372
QY 2011 ATGCTTATGTGCAACATTTAATGAATATATGAAACCGCTTACGTGCTATTTGCAATTAG 2070
Db 1373 GGCATTTGTACCAACAGTTGAGGAATACATGSCAAATGTCAGTTTCTCTGCGCACTGG 1432
QY 2071 GCCCGATTGTCAAGCGGCTATTTTCTTGTGGGCGCCAAATTTATCAGAGGAGATTTGTTG 2130
Db 1433 CGGTCAATTAATCTTCCAGCGCATGATTTTCTTGGCGAAACGCTCTCAGATTAATGTTCA 1492
QY 2131 AAAGCTCTGAATATCATAATCTATTTAAGCTAATGAGCAGCGAGGTCGACTTTCTTAAACG 2190
Db 1493 AAGATCATGAATACAGCAATTTCTGAGCTCATGTTCTTCTCAGTCTGCTCTTAAATG 1552
QY 2191 ATATCCATAGCTTCAAGAGGAAATTTAAGAAAGCAATTTAAACGCGGTAGCATTTGCAATT 2250
Db 1553 ACATTCGAACGTTGAGAGGAGTTTCAAGCTGTTAACTGAAACGCGTTTCTACTGCTTG 1612
QY 2251 TGAGTAAACGGAAGAGTGGGAAAGTGGAAAGAGGTTGTGGAGGAGATGATGATGATGA 2310
Db 1613 CTCTACAGCTGGTGTTCATGTCCTATAGAGCGCTAAGAGGAACACATGATCTT 1672
QY 2311 TTAAGAAACAGAGGAAGAAATTAATGAATTAATTTTGAAGAAATTTGAGCAATTTGTTTC 2370
Db 1673 TAGAGTCGTATAGGAGAGGTTTAGTAGCTGTGGTTGCT---AGACAGGACAGATTTGTTTC 1729
QY 2371 CTAGAGCTTGAAGATGCAATTTTGGACATGTTGACGTTGCAATTTTATTTTATTCGCAA 2430
Db 1730 CTAGGTCAATCAAGGAGCTGTTCTGGAAGCTTTTGAAGCAATTTCACTGTTCTACTTCC 1789
QY 2431 ACGATGACGGGTTTAC 2446
Db 1790 AGATTGATGATTTAC 1805

RESULT 9
US-10-425-114-17004
; Sequence 17004, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 17004
; LENGTH: 2086
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3067-045-E3_FLI
US-10-425-114-17004
Query Match 15.5%; Score 433.2; DB 16; Length 2086;
Best Local Similarity 55.8%; Pred. No. 2.7e-91;
Matches 947; Conservative 0; Mismatches 713; Indels 36; Gaps 5;
QY 778 ATTCAATGAGATGAGATGATCTGCGTATATCTCTGAAGGACTCGGTAAATTTATATG 837
Db 119 AATCTACGCAATAGAGCATATTTGCCATGTTCTGAGAGGTTGGAACCTGCTGG 178
QY 838 ATTGGAATATGTTGAAGAAATATCAGATGAAAAATGTTCTGTTTCAACTCACCATCAG 897
Db 179 ACTGGAATGATGTTATGAAGTTCCAAAGCGAAGATGGATCCTGTTTAACTCTCCTCTG 238
QY 898 CAAAGCTGCTGTTTCAATTAATCATCAAAATCCTGTTGCTTAATTTAAATTTAC 957
Db 239 CAACTGCTGCGCTTTGGTGGCCAACTATGACGACAAAGCGCTACAGTATCTAAATTTGC 298
QY 958 TTTTGGACAGTTTGTATGACAGTCCCAACAGTTTATCTCATGATTTATTTATCCGAC 1017
Db 299 TTGTCACAAATTTGGCAGTGCAGTACCAACAGTTTCCCAAAATTTCACTATCAGC 358
QY 1018 TTTCTATGTTGACAAATTTGAAGATTTAGGAATTTTCAACCAATTTTCAAGTGGAAATTA 1077
Db 359 TTTCAATGTTGACACGCTCGAAAGTTTGAATATCAAGGCAATTTTCTGTGGAGAAA 418
QY 1078 AAAATGTTTATGATGAAACATACAGATGTTGGTGGACAGATGACCAATTTATCATGG 1137
Db 419 AGGCTGTCTCGACATGATATACAGTTTGTGGTTACAGAAAGAGGAGGAATTAATGTTG 478
QY 1138 ATGTTGTAACATGCTTTTACGCTTTTCGGTTTATTAAGCATCAATGGGTATGAAGTTTCCC 1197
Db 479 ATGCAAAACATGTCGATGCGAATTTTCAATTTTCAATGAATGCTTTGATGTTTCCA 538
QY 1198 CAGATCCATTTGCTGAAATTTACTAATG-----AATAGCTT 1233
Db 539 CAGATTGTTGTTCTCATATTGCTGAAGCTTCCAAATTTCCAACTCACTCAAGGATATT 598
QY 1234 TGAAGACGATATGAGCTCTTGAACATATCATGCTGTC---ACATATATTATACCAAG 1290
Db 599 TATGTGATCAAGACCTCTATTGGAATGATGACAAAGCTCTTAAAGTCTATATTATCAGAAA 658
QY 1291 AGGATTTATCTTCTGGAACAAATCTTGAAGTCAAGTCAAGTCTGATTTTCTCAAGAGATAATAT 1350
Db 659 GAGACTTCACTTATGAGAACATAGATGTTGACAGGTAGCTTATTTAGGGAAAGCTAT 718
QY 1351 CCAGTGTATCAACAGGCTTTCTTAATTAATTTCAAGAGGTGGAATAATGCTCTTAAGT 1410

RESULT 10

US-10-425-114-5988

; Sequence 5988, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; NUMBER OF SEQ ID NOS: 2003-04-28

; SEQ ID NO 5988

; LENGTH: 2029

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURES:

; OTHER INFORMATION: Clone ID: 700550363_FLI

US-10-425-114-5988

Query Match 14.1%; Score 392.4; DB 16; Length 2029;

Best Local Similarity 53.9%; Pred. No. 1.1e-81;

Matches 999; Conservative 0; Mismatches 741; Indels 112; Gaps 5;

Qy 700 TGGACATAAACTCTCTTTCAAAACAAACAGATTTTGTGATGCTACATAAGAGGAAT 759
 Db 4 TGGGTTAGATTTCCTGTAAGACAACTGATGCTGGCATCTTCCCGCCGGGAGA 63

Qy 760 TGGAGCAAAAAG-----ATGCCAATCAAAATGAGATGGATGATTTGGCGTATA 810
 Db 64 TGGAAATGAAAAGGCTGGCTGGAGTAGTCTCTTTTGGAAAGAAAAGCATATATGGCTTTA 123

Qy 811 TCTCTGAGGACTCGGTAATTTATATGATTTGATGATGCTGTAAGAAATATCAGATGAAAA 870
 Db 124 TCCAGAGGAAATCGGAATATGCTGGACTGGATCAAGTTATGAAGTTTCAGAGGAGA 183

Qy 871 ATGTTCTGTTTCACTCACCATCAGCAACAGCTGCTGCTTTCAATTAATCATCAAAATC 930
 Db 184 ATGATCATTTCTCAGCACTCTTCCCACTGCTGTTGCAATTAATCCAAATACACAG 243

Qy 931 CTGGTTGCTTTAATTTAATTAATCACTTTTGGACAACTTTGGTAATGAGTCCCAACAG 990
 Db 244 ACCAAGCCCTTCAATACCTAAATTCGTTGTCAGTGAATTTGGCAGTGCAGTACCAGCA 303

Qy 991 TTTATCCTCATGATTTATTTATCCGACTTTCTATGGTTGACACAAATGGAAGATTAGGAA 1050
 Db 304 TGTATCCTTCAATGGTACATTTGTCAGCTTTTCAATGGTGGACGCGCTTGAAAAAATGGGAA 363

Qy 1051 TTTTCAACCACTTTTCCAGAGTGAAATTTAAATATTTTATAGTGAACATACAGATTTGGG 1110
 Db 364 TTTCTCAGCGCTTTGTCAGTGAATTTAGAAAGCATCTTGGACATGGCATCAATTTGCTGGT 423

Qy 1111 TGGAAACAGATGAGCAATATTTCAATGATGTTTGAACATGCTTTTAGCCCTTTCCGGTTAT 1170
 Db 424 TACAGAAAGATGAGAAATCATGATGGACATAGCAACATGTCATGGCAATTTCCGCTTT 483

Qy 1171 TAAGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTAATGAATAG 1230
 Db 484 TGAGATGAATGGGTATGAATTTTCCCTCAGATGAGTGTCTCAGCTTGTGGAGCTTCCA 543

Qy 1231 CTTTGAAGACGAATATGCACTCTTGAAA----- 1260
 Db 544 CTTTCCATGATTCATCAGGAATTTTAAATGATACAAATCCCTACTGGAATTTGACA 603

Qy 1261 -----CATATCATCGTCAACATATATTATACCA 1288

Db 604 AGACCTCAAAAGTCACCTTATCAGAAAAAGATCTGATCTTAGATCGCATAGGTTCTCTGGT 663
 Qy 1289 AGAGGATTTATCTTCTGGAACAAATCTTGAAGTCAGCTGATTTCTCTCAAGAGATAT 1348
 Db 664 CTGGCAACTTATTTGAAGGATAGATGCTGCTAGTAGGGTGCAGAAAGACTCGATTTTGG 723
 Qy 1349 ATCCACTGATTTCAACAGGCTTTCTAAATTAATTCACA----- 1386
 Db 724 GAGAGGTCTGCAGAAAAAATTTAAATTTCTATTTTCACCTGGGAGGTTCAAGTTATGTC 783
 Qy 1387 -----AAGAGTGGAAAAATGCTCTTAAAGTTCCCTATCAATACCGTTTAGAA 1433
 Db 784 TGTGTTTTGTGATTTTCAAGTGGATGCTGCTTAATTTTCCCTTGTATTTCCACATGGAG 843
 Qy 1434 CGCATAAACACTAGACGAAATATACAGCTTTACAAATGAGACAAATACAGAAATTCGAAA 1493
 Db 844 CGTCTAGAACACAAGAGAAACATCGAACATTTTGTATGCTGGGGTTCTCTGATGCT--A 900
 Qy 1494 ACTACATATCATCATCAATATTTAGTAACACTGATTTACCTTAAGTTGGCTGTTGAAGAT 1553
 Db 901 AACAAAAATCCTCATCTTTTCGTATCAATCAAGAAATTCCTAGCTTTGGCAGTCAGAT 960
 Qy 1554 TTCTACACCTGCCAATCTATTTATCGTCAAGAAATTTAAAGGTTCTTGAAGTGGGTGTA 1613
 Db 961 TTCAGTTTCTCTCAACGTTTACCGGATGAATTCGGCATCTTGATAGTTGGGTGAAG 1020
 Qy 1614 GAGATTAAGTTGGACCGAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCT 1673
 Db 1021 GAGAACAGCTGGACCGAGCTACAATTTGCTCGGAGAAATCGACATATTCGTATCTGCT 1080
 Qy 1674 GTTCTGCAACACTTTGCTCTCCGAATATCAGATGCGCTTATTCATGGGCCCAAAAT 1733
 Db 1081 GCTGCTGCTACCGTATTTTCTCTGAATTTGTCACGCTCGCATTTCAATGGGCCAAAAAT 1140
 Qy 1734 GGCATATTTAACTACAGTAGTTGATGACATTTTTTCAATTCGGTGTGTAATCGATGAATG 1793
 Db 1141 GGTGCTCTCAACTGTGTTGATGATCTTCTCGATGTTGTTGGATCAAAGAGAAATTA 1200
 Qy 1794 ACCAACCTGATCAATGTTTGAAGAAATGGAATGAGATGCGACAGGAATGTTGTTCA 1853
 Db 1201 GAAAACTGTATGACTAGTTGAGAAATGGCATCGCACCATGCGATGAGTTGATTTCTATTCG 1260
 Qy 1854 GAGCATGTTGCGATTTTATTTTATAGCATTTAAAGATGCAATCTGTTGATTTGGAGATCA 1913
 Db 1261 GAACAGGTGAATATGATTTTCTGCTATTTATACACAGTGAACCATCTTTGGAGCAATG 1320
 Qy 1914 GCTTTTAAATGGCAAGCGCGGATGTAACTAGCCATGTTATTTCAAACTTGGTGGAACTA 1973
 Db 1321 GCTTCTGCAGCACAAAGCCGCTGATCTTTACAAACCACCTAGTAGAAATATGCTGGAATTTG 1380
 Qy 1974 ATGAATAGTATTTGAGAGAGCTATATGAGCAAGAGATGCTTATGTCGCAACATTAAT 2033
 Db 1381 TTAAGATCTATGATGTTGAGGCAAGATGCGAGAGATGCCAATATATACCAACAGTTGAA 1440
 Qy 2034 GAATATATGAAAAAGCTTACGCTCATTTGCAATTTAGCCCGATTTGTCAGCCGCTATT 2093
 Db 1441 GAATACATGACAAATGCTGTTGCTCATTTGCACTGGGCGCAATTTGCTCCGAGCATG 1500
 Qy 2094 TACTTTTGGGGCCCAAAATATCAGAGAGATTTGTTGAAGACTCTGAATATCAATATCTA 2153
 Db 1501 TATTTTGTAGGGCAAGAGCTATTAGAGCATGCTGTCTAAAGATGAGAGTACGATTAATTA 1560
 Qy 2154 TTTAAGCTAATGAGCAGCGAGGTCGACTTCTTAAACGATATCCATAGCTTCAAGAGGAA 2213
 Db 1561 TTTAGGCTAGTAGCACTTTCGGGAGGCTCTCTAATGACTACCAAGATTTTAGAGAGGAA 1620
 Qy 2214 TTTAAGCAAGCAAAATTAACCGGTAGCATTTGATTTGATTAACGAGAAAGTGGGAA 2273
 Db 1621 GGCACAGGGAGCTGAATAGTGTCTTCTACTTGTCTCCACAGTGGTGTCTTATG 1680
 Qy 2274 GTGAAAGAGAGGTTTGGAGGAGATGATGATGATTAATAAACAAGAGAGAAAGATTA 2333
 Db 1681 TCCATAGAACCCGCTAAAGAGGCAATGCAAGAGTCCATAGACGCTGTCTAGAGAGACTTG 1740

QY 2334 ATGAAATTAAATTTTGAAGAAATGCTAGCATTTCTTCTAGAGCTTGTGAAAGATGCAATTT 2393
Db 1741 CTAAGATTGGTCT---CAGGAAGAAGTGTCTTCTTAGGCCAAGCAAGAGCTCTTC 1797
QY 2394 TGGAACTGTGTGACGTGTGTGAATTTTTTTTACGCAAAAGATGACGGGTTTA 2445
Db 1798 TGGAGATGTGTAAAGATCTTACCTGTTTCTCTCAGATGATGATTTTA 1849

RESULT 11
US-10-259-194A-107
; Sequence 107, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassenian, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 107
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-107

Query Match 13.6%; Score 379.4; DB 16; Length 2223;
Best Local Similarity 54.5%; Pred. No. 1.3e-78;
Matches 900; Conservative 0; Mismatches 676; Indels 75; Gaps 4;

QY 795 GGATACCTTGGCGTATATCTCTGAAGACTCGGTAAATTTATATGATTTGGAATATGTTGAAG 854
Db 574 GAATATACGGCTTAUTCTCTCAGAGGATTTAGGCAATTTTCAGNACTGGATGAGTGAAG 633
QY 855 AAATATCAGATGAAATGGTCTCTTTCACTCAACATCCAGCAACAGCTGCTGCTTTC 914
Db 634 AAATCCAAAGAAAGATGGCTCACTGTTCACCTCCCTTACACAACTGCAGCTGCATTA 693
QY 915 ATTATCATCAAAATCTGGTGTCTTAATTTAAATTTCACTTTTGGACAGTTTGGT 974
Db 694 GTCCCAACATATGATGCCAAAGCTCTCCAGTACTTAGACATGTTCTTGGACAAATTTGA 753
QY 975 AATGAGTCCCAACAGTTTATCTCATGATTTATTTATTTATTCAGTCTTTCTATGTTGACACA 1034
Db 754 AGTGAGTCCAGCGGCTTATCTTGAATTTTCAGTCTCAGCTCTACATGTTGGATGTG 813
QY 1035 ATTGAAGATAGGAATTTCAACCATTTTCAGAGTGGAAATTTAAATGTTTTAGATGAA 1094
Db 814 CTTGAAAGATGGGAATATCTAGGCATTTTGTGTGAGATAAGAGCATCTGGACATG 873
QY 1095 ACATACAGATGTTGGTGGAAACGAGATGAGCAATTTATCATGATGTTGTAAATGCT 1154
Db 874 ACCTACAGTTGCTGGAAACAGAGGATGAGAAATTTGCTTGACATGCAATGTTGG 933

| Query Match | 11.8%; | Score 330; | DB 16; | Length 1559; |
|-----------------------|-----------------|--|-----------|--------------|
| Best Local Similarity | 55.7%; | Pred. No. 4.9e-67; | | |
| Matches 655; | Conservative 0; | Mismatches 515; | Indels 6; | Gaps 1; |
| QY | 1270 | CGTCACATATATTATACCAAGAGATTTATCTTCGAAAAACAAATCTTGAAGTCAGCTG | 1329 | |
| Db | 116 | CGACAGTTAGTATCTCTGAAGTAGTCTATCTCGGATAGCNVAGGCTCAGGTCACGTA | 175 | |
| QY | 1330 | ATTTCTCTCAAAAGAGATAATPCCACTGATTCAAACAGGCTTTCTAAATTAATTCACAAAG | 1389 | |
| Db | 176 | CTTTACTGAGGGAAACAACCTAGAGTCTGGTGGTCTCTACGAAAAACCTTCACCTCTTTAAAG | 235 | |
| QY | 1390 | AGGTGGAAAAGTCCTTTAAGTTCCTATCAATACCGGTTTAGACGCATAAACAACATAGAC | 1449 | |
| Db | 236 | AGGTGGAAACATGCTCTGGACCGCTCTCTACACCAACATTGGACCGCTCACACCATAGGT | 295 | |
| QY | 1450 | GAATAATACAGCTTTTACAATGTAGACAATAACAAGATTTCTGAAAACATACATATCACTCAT | 1509 | |
| Db | 296 | GGACATCGAAAATTTCAATATTATAGACGAGCACATGCTAGAGACACCATACTTGTCAA | 355 | |
| QY | 1510 | CAAAATATTAGTAAACACTGATTAACCTTAGTTGGCTGTGTGAAGATTTCTACACCTGCCAAT | 1569 | |
| Db | 356 | ATCAACATACCAGTAGAGATATTTCTAGCGTTTGAGTATTAGAGACTTCAGTTCTCTCAGT | 415 | |
| QY | 1570 | CTATTTATCGTGAAGAAATTTAAAGTCTTTGAAAGCGTGGTAGAGAAATGATTGTGACC | 1629 | |
| Db | 416 | TTACTTACAGCAAGACTTCAACATCTTGAAGACTGGTGAAGAGTGCAGGTTAGACC | 475 | |
| QY | 1630 | AGCTCAAGTTTGCTAGGCAAAAGACCGCCTACTGTTATTTCTCTGTTGCTGCAACACTTT | 1689 | |
| Db | 476 | AGCTACAAATTTGGCGACAGAAAGTTGGCATACTTCTACTCTGCTGCTGGCACCATGT | 535 | |
| QY | 1690 | CGTCTCCGAATTATCAGATGGCGGTATTTTCATGGCCAAAAATGGCATATTAACTACAG | 1749 | |
| Db | 536 | TCTCTCTTGAGTGCTGATGCTCGAACTTTGTGGGGCAAAAATGGTGTGCTTCAACATA | 595 | |
| QY | 1750 | TAGTTGATGACTTTTTTTGATATPCGGTGGTGTACAATCGATGAATTGACCAACTGATTCAAAT | 1809 | |

```

RESULT 14
US-10-437-963-72152
; Sequence 72152, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 72152
; LENGTH: 3666
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72557C.1
US-10-437-963-72152

```

| | | | | |
|-----------------------|--------------|--|--|-------------------|
| Query Match | 11.3%; | Score 315.6; | DB 17; | Length 3666; |
| Best Local Similarity | 58.1%; | Pred. No. 1.9e-63; | | |
| Matches 615; | Conservative | 0; | Mismatches 434; | Indels 9; Gaps 3; |
| QY | 1388 | AGAGGTGGAATGCTCTTAGTGTCCCTATCAATACCGGTTTGAACGCGATAAACA | CTAG | 1447 |
| DB | 2325 | AAAATGAAGTATGCTCTTAAGTTCCCTTCTACACCACTAGGACCGCTTAGATCAAA | 2384 | |
| QY | 1448 | ACGAAATATACAGCTTTTACAATGTAGACAAATACAAGAAATCTGAAAACTACATATCACTC | 1507 | |
| DB | 2395 | GAGGAACATTGAACGTTTTGATGCAAGGATTC | CCAGATGTTAAAGACGGAT---ACTT | 2441 |
| QY | 1508 | ATCAAAATATAGTACACTGAATACCTTAAGGTGGCTGTTGAAGATTTCTACACCTGCCA | 1567 | |
| DB | 2442 | GCTTCCTCATGCCAATCAAGATATCTAGCTTTGGCTTTGAAGATTT | CAGTAGTTCTCA | 2501 |
| QY | 1568 | ATCTATTTATCGTCAAGAAATTAAGGTCCTTGAAGGTGGTGGTGTAGAAATTAAGTTGGA | 1627 | |
| DB | 2502 | ATCTATATACCAGGATGAACATTAATATCTTGAGTGTGGGTGAAAGATGAAAGCTCGA | 2561 | |
| QY | 1628 | CCAGCTCAAGTTTCTTAGGCAAAAGACCGCTACTGTTATTTCTCTCTGCTGCAACACT | 1687 | |
| DB | 2562 | TCAGTGCCATTUGCA | CGCAGAGTTGACATATGCTACTCTTCTCTGCTGCTACCAT | 2621 |
| QY | 1688 | TTGCTCTCCGAAATATCAGATGCGCGTATTTTCATGGGCCCAAAATAGCCATATTAACATC | 1747 | |
| DB | 2622 | ATTTCCCGGTGAATGTGCTGAAGCCCGCATTTGCATGGCTTAAATATGTTGCTAGCAAC | 2681 | |
| QY | 1748 | AGTAGTTGATGACCTTTTTCATGATCGGTGGTACAAATCGATGAATGACCAACCTGATTC | 1807 | |
| DB | 2682 | TGTTGTTGATGACCTCTTTGATCTTTGGGGGATCAAAAGAAAGAACTAGAAACCTCATTCG | 2741 | |
| QY | 1808 | ATGTGTTGAAAAATGGAATCTAGATGTCGCAAGGATTTGTTGTTTCAGAGCATGTCGGAT | 1867 | |
| DB | 2742 | TTTAGTTGAGAGTGGGATGGACAT---CAAGAGGAGTTCCTCTCAGAACAGATAGAAT | 2798 | |
| QY | 1868 | TTTTATTTTATAGCAATAAAGATGCAATCTGTTGGATTGGAGATCAAGCTTTTAAATGGCA | 1927 | |
| DB | 2799 | AGTTTTTCTGCTATTTATCTACAGTGAACCGAGCTTTGGAGCAAGGCTTCTGCATTACA | 2858 | |
| QY | 1928 | AGCGCGGATGTAACTAGCCATGTTATTCAAACTTGGTTGGAACTTAAGTAATAGTATGTT | 1987 | |
| DB | 2859 | AGGCGGTGATGTTACAAAAACCCCTAACAGAAAAATATGGTTATGTCTCATGAGGCTATGAT | 2918 | |
| QY | 1988 | GAGAGAAGCTATATGGAACAAGATGCTTATGTGCCAACATTAATAAGTATATATGGAAA | 2047 | |
| DB | 2919 | GACCGAGCTGATGGCAGAGGCAAAATATGTGCCACAAATGGAAGAATATATGCAAA | 2978 | |
| QY | 2048 | CGCTTAGTGTCAATTTGCAATTAGCCCGCATTTGTCAAGCCGGCTATTTACTTTGTGGGGC | 2107 | |
| DB | 2979 | TGCTGTGTTCTCATTTGCACTGGGACCTATTTGTGCTCCGACTCTGTATTTCTAGGACC | 3038 | |
| QY | 2108 | CAAAATATCAGAGGAGATTTGTGAAAGCTCTGAATATCATATATCTATTTAAGCTAATGAG | 2167 | |
| DB | 3039 | GAAGCTCCAGAGGATGTCTGTGAGGGATCATGAGTACAATGAATATTTAGACTGATGAG | 3098 | |
| QY | 2168 | CACGCGGGTCGACTTCTTAAACGATATCCATAGCTTTCAGAGGGGAATTTAAGGAGGCA | 2227 | |
| DB | 3099 | CACTTGTGGCGTCTCTGATGACAGCCAGGCTTTGAGAGGGAGAGCCTGGAGGGAAA | 3158 | |
| QY | 2228 | ATTTAAACGGGTAGCATTTGATTTGAGTACCGAGAAAGTGGGAAAGTGGAAAGAGGT | 2287 | |
| DB | 3159 | GCTGAACAGTGTCTCACTGCTGTTTCATCACAGTGGTGTCTATCTCCATAGACGAGGC | 3218 | |
| QY | 2288 | TGTGGAGGAGATGATCATGATGATTAATAAACAAGAGAAAGAAATTAATGAAATTAATTTT | 2347 | |
| DB | 3219 | TAAATTAAGAACCCAGAAAAATCCATAGACATTTCCAGGAGGAACCTTCTTAAGATTGGTCT | 3278 | |
| QY | 2348 | TGAAGAAAAATGGTAGCATTTGTTCTTAGAGCTTGTAAGATGCAATTTTGAAATGTCAT | 2407 | |
| DB | 3279 | TGAGAAACAGGHT---GCTGTTCTAGCCATCAAGCAGCTGTTCTGGAAGATGTGCAA | 3335 | |

| | | | |
|--|------|---|------|
| Qy | 2408 | CGTGTGTAATTTTTTTACGCAACGATGACGGGTAA | 2445 |
| | | | |
| Db | 3336 | GATTGTTACATGTTCTACTCTCGGACTGATGGGTAA | 3373 |
| | | | |
| RESULT 15 | | | |
| US-10-425-114-729 | | | |
| ; Sequence 729, Application US/10425114 | | | |
| ; Publication No. US20040034888A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Liu, Jingdong | | | |
| ; APPLICANT: Zhou, Yihua | | | |
| ; APPLICANT: Kovalic, David K. | | | |
| ; APPLICANT: Screen, Steven E | | | |
| ; APPLICANT: Tabaska, Jack E | | | |
| ; APPLICANT: Cao, Yongwei | | | |
| ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With | | | |
| ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement | | | |
| ; FILE REFERENCE: 38-21(53313)B | | | |
| ; CURRENT APPLICATION NUMBER: US/10/425,114 | | | |
| ; CURRENT FILING DATE: 2003-04-28 | | | |
| ; NUMBER OF SEQ ID NOS: 73128 | | | |
| ; SEQ ID NO 729 | | | |
| ; LENGTH: 1150 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Zea mays | | | |
| ; FEATURE: | | | |
| ; OTHER INFORMATION: Clone ID: 700088295_FLI | | | |
| US-10-425-114-729 | | | |
| Query Match 10.3%; Score 288.8; DB 16; Length 1150; | | | |
| Best Local Similarity 59.1%; Pred. No. 2.1e-57; | | | |
| Matches 513; Conservative 0; Mismatches 352; Indels 3; Gaps 1; | | | |
| Qy | 1578 | CGTGACGAATTTAAAGTCTTTGAAGTGGTGGTAGAGAAATAGTTGGACCAAGCTCAAG | 1637 |
| Db | 3 | CGGATGAACTTCGGCACTCTTGATAGTTGGTGAAGGAGAACCAAGCTGGACCAAGCTACAA | 62 |
| Qy | 1638 | TTTGCTAGGCAAAAGACCGCTACTGTATTCTCTCTGTTGCTGCAACACTTTCGTCTCCC | 1697 |
| Db | 63 | TTTGCTCGGCAGAAACCTGACATATTGCTATCTCTGCTGCTGCTACCGTATTTTCTTCT | 122 |
| Qy | 1698 | GAATTAFCAGATCGCGGTATTTTCATGGGCCAAAATGGCATATTAACACAGTAGTTGAT | 1757 |
| Db | 123 | GAATTGCTGCACCTCGCAATTCATGGGCCAAAATGGTGTCTCTCACACTGTGGTTGAT | 182 |
| Qy | 1758 | GACTTTTTTGATATCGTGGTACAAATCGATGAATGACCAACCTGATTCAAATGTTGAA | 1817 |
| Db | 183 | GACTCTTCGATGTTGGTGGATCAAAAGAAGAAATTAGAAAACCTGTAGCACTAGTTGAG | 242 |
| Qy | 1818 | AAATGGAATGTAGATGTGCACAAGGATGTGTTTCAGAGCATGTTCCGGATTTTATTTTAA | 1877 |
| Db | 243 | AAATGGCATGGCCACCATGCAGTTGAGTTCTATTTCGAAACAGGTGAAATAGTATTTTCT | 302 |
| Qy | 1878 | GCATTTAAAGATGCAATCTGTTGGATTTGAGATGAAGCTTTTAAATGGCAAGCGCGCAT | 1937 |
| Db | 303 | GCTATTTATACAACAGTGAACCATCTTTGAGCAATGGCTTCTCGACCAAGAAGCGTGTAT | 362 |
| Qy | 1938 | GTAACTAGCAGTTATTTCAAACCTGTTGGAACTTAATGAATAGTATGTTTGAGAGAAGCT | 1997 |
| Db | 363 | CTTACAAAACCACTAGTAGAAATATGGCTGSGATTTTGTAAAGATCTATCATGTCGAGGCA | 422 |
| Qy | 1998 | ATATGACACAAGAGATGCTTATGTGCCAACATTAATGAATATATGGAACCTTACGTG | 2057 |
| Db | 423 | GAATGGCAGAGATGCCAATATGTACCAACAGTTGAAGAATAATGACAAATGCTGTGTC | 482 |
| Qy | 2058 | TCAATTTGCATTAGGCCCGAATGTCAAGCCGGCTATTTTACTTTGTGGGGCCAAATATCA | 2117 |
| Db | 483 | TCAATTTGCCATGGGCCCAATTTGCTCCAGCATTTGTAATTTTGTAGGGCAAGAGCTATTA | 542 |
| Qy | 2118 | GAGGAGATTTGTAAGCTCTGAAATATCATATCTATTTAAAGCTAATGAGCAACGAGGT | 2177 |
| Db | 543 | GAGCATGCTGTCAAAGATGAAGAGTACATAAATTTATTTAGCTAGTGAACACTTGGCGG | 602 |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:43:59 ; Search time 228.559 Seconds
(without alignments)
8682.774 Million cell updates/sec

Title: US-10-041-018-361

Perfect score: 2792

Sequence: 1 cccactcatctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----|-------------------|
| 1 | 782.2 | 28.0 | 2658 | 3 | US-08-727-308-2 |
| 2 | 782.2 | 28.0 | 2658 | 3 | US-08-727-308-3 |
| 3 | 453.2 | 16.2 | 2403 | 4 | US-09-614-912-23 |
| 4 | 238.2 | 8.5 | 1157 | 4 | US-09-614-912-33 |
| 5 | 194.8 | 7.0 | 2861 | 4 | US-09-398-395A-55 |
| 6 | 194.8 | 7.0 | 2861 | 4 | US-09-887-586A-55 |
| 7 | 194.8 | 7.0 | 2861 | 4 | US-09-895-752-55 |
| 8 | 194.8 | 7.0 | 2861 | 4 | US-09-903-012B-55 |
| 9 | 194.8 | 7.0 | 2861 | 4 | US-09-900-797-55 |
| 10 | 153.8 | 5.5 | 2700 | 3 | US-09-315-861-1 |
| 11 | 153.8 | 5.5 | 2700 | 4 | US-09-398-395A-43 |
| 12 | 153.8 | 5.5 | 2700 | 4 | US-09-887-586A-43 |
| 13 | 153.8 | 5.5 | 2700 | 4 | US-09-895-752-43 |
| 14 | 153.8 | 5.5 | 2700 | 4 | US-09-903-012B-43 |
| 15 | 153.8 | 5.5 | 2700 | 4 | US-09-593-253-1 |
| 16 | 153.8 | 5.5 | 2700 | 4 | US-09-900-797-43 |
| 17 | 139.2 | 5.0 | 531 | 4 | US-09-614-912-27 |
| 18 | 130.6 | 4.7 | 468 | 4 | US-09-614-912-25 |
| 19 | 120.8 | 4.3 | 2528 | 3 | US-09-234-393-41 |
| 20 | 120.8 | 4.3 | 2528 | 4 | US-09-865-171-41 |
| 21 | 119.2 | 4.3 | 2424 | 3 | US-09-234-393-1 |
| 22 | 119.2 | 4.3 | 2424 | 4 | US-09-360-545-15 |
| 23 | 119.2 | 4.3 | 2424 | 4 | US-09-865-171-1 |
| 24 | 119.2 | 4.3 | 2424 | 4 | US-09-398-395A-45 |
| 25 | 119.2 | 4.3 | 2424 | 4 | US-09-887-586A-45 |
| 26 | 119.2 | 4.3 | 2424 | 4 | US-09-895-752-45 |
| 27 | 119.2 | 4.3 | 2424 | 4 | US-09-903-012B-45 |

28 119.2 4.3 2424 4 US-09-900-797-45 Sequence 45, Appl
29 119.2 4.3 2525 3 US-09-234-393-39 Sequence 39, Appl
30 119.2 4.3 2525 4 US-09-865-171-39 Sequence 39, Appl
31 119.2 4.3 2528 3 US-09-234-393-37 Sequence 37, Appl
32 119.2 4.3 2528 4 US-09-865-171-37 Sequence 37, Appl
33 119.2 4.3 2571 3 US-09-234-393-12 Sequence 12, Appl
34 119.2 4.3 2571 4 US-09-865-171-12 Sequence 12, Appl
35 113.4 4.1 2460 4 US-09-371-307-2 Sequence 2, Appl
36 100.6 3.6 2543 4 US-09-614-912-21 Sequence 21, Appl
37 90.4 3.2 1865 4 US-09-398-395A-47 Sequence 47, Appl
38 90.4 3.2 1865 4 US-09-887-586A-47 Sequence 47, Appl
39 90.4 3.2 1865 4 US-09-895-752-47 Sequence 47, Appl
40 90.4 3.2 1865 4 US-09-903-012B-47 Sequence 47, Appl
41 90.4 3.2 1865 4 US-09-900-797-47 Sequence 47, Appl
42 90.4 3.2 1885 3 US-09-234-393-45 Sequence 45, Appl
43 90.4 3.2 1885 3 US-09-234-393-45 Sequence 45, Appl
44 90.4 3.2 1885 3 US-09-234-393-47 Sequence 47, Appl
45 90.4 3.2 1885 4 US-09-865-171-43 Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-08-727-308-2
; Sequence 2, Application US/08727308
; Patent No. 6020176
; GENERAL INFORMATION:
; APPLICANT: Yuji KAMIYA et al.
; TITLE OF INVENTION: KAURENE SYNTHASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,308
; FILING DATE: October 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-727-308-2

Query Match 28.0%; Score 782.2; DB 3; Length 2658;
Best Local Similarity 61.3%; Pred. No. 1.1e-191;
Matches 1390; Conservative 0; Mismatches 823; Indels 54; Gaps 6;

QY 275 AACTATCCCACTAACTGATCATGTATACACCAAGACGGATCCAAAACAGTTAA 334

| | | | |
|----|------|---|------|
| Db | 236 | AACTAAABCTGGAGCTTTGCACTTTGAAGAAACAAAAGAAAGAAATTAAGAAATTTGTTGCA | 295 |
| Qy | 335 | AAATGTAGAAATTTCTGTTTCTTCAATATGACACAGCATGGGTAGCCATGGTCCCTCTCC | 394 |
| Db | 296 | CAAGGTTGAACCTTTCAGTTCTGCAATATGATACACTGATGGGTGGCAATGGTTCTCTCC | 355 |
| Qy | 395 | AAACTCACCAATCGCCTTGTTCCTTGAGTGTCTCAATTTGGTTAAATTAATAACAGCT | 454 |
| Db | 356 | AAACTCTCTCAACCAACCTCTTTTCCCGAGTGATATAAAGTGGTATAGATAGTCAACA | 415 |
| Qy | 455 | TAATGATGGTTCATGGGTCCTTTGTTAAATCACACTCAATATCAATCAATCAACCGCTTGTAA | 514 |
| Db | 416 | TGCTGATGGCTCATGGGCTACTCCACAAC-----GATCAGTTGCTGATGAA | 463 |
| Qy | 515 | AGATCTCTATCTTCAACATTAGCATGATTTGTTGCATTAATAAGATGAAATTTGGGGA | 574 |
| Db | 464 | GGCCATCTCTTATCTACATATGACATGTGTTCTTACTCTTAAGCGGTGGAATATTGGGCA | 523 |
| Qy | 575 | AGATCAAAATAAATAAGGCTTAAGTTTATTCAGTCAAAATCTTGCTTCAGCTACTGAAAA | 634 |
| Db | 524 | TGATCATATGACCAAGGCCCTTGATTTATCAAGTCTAATATAGCTTCAGCTACTGATGA | 583 |
| Qy | 635 | AAAGTCAACCATCTCCATTTGTTTGAATCATATTTCTGTTGCTTGGTTGCTTGGATGCGAA | 694 |
| Db | 584 | GAAACCAACGTTCTCGGTGGGATTTGACATTAATTTTCCCTGGCATGATGAGTATGCTAA | 643 |
| Qy | 695 | AAACTTGGACATAAACCCTCTTTCAAAAACAAACAGATTTTACGTTGATGCTACATAAGAG | 754 |
| Db | 644 | AGACTTGAATTTGAATCTACCTTGGCACCGAGAACGTTGGATGCTTGGTTGCAAGAA | 703 |
| Qy | 755 | GGAATTTGGAGCAAAAA-----AGATGCCATTCMAATGAGATGGATGATCTTGGCGTA | 808 |
| Db | 704 | AGAGTTGGAGCTGAGAGCTGCAGAGCAACTCTGAGGTGGAAAGCCATTTACGCTA | 763 |
| Qy | 809 | TATCTCTGAAGGACTCGGTAATTTATGATTTGGAATATGGTGAAGAAATATCAGATGAA | 868 |
| Db | 764 | TGTTTCAGAAAGAAATGGAAGTTACAGGACTGGGATATGGTCATGCAATATCAAAAGAA | 823 |
| Qy | 869 | AAATGTTCTGTTTTCAACTCACCATCAGCAACAGCTGCTGCTTTCATTAATCATCAAAA | 928 |
| Db | 824 | GAAATGATCACTGTTTAAATCTCCATCCACTACGGCAGCGGCTTTTATGCAATAGAATGA | 883 |
| Qy | 929 | TCTGCTGTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 988 |
| Db | 884 | TGATGCTGTTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 943 |
| Qy | 989 | AGTTATCTCTATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 1048 |
| Db | 944 | AATATATCTCTTGATATATATGCTCGATTAACATGATGGTTGATAGCTTCAAAAATTCGG | 1003 |
| Qy | 1049 | AAATTTACACCATTTTCAGAGTGAATTAATAATTAATTAATTAATTAATTAATTAATTAAT | 1108 |
| Db | 1004 | AAATTTCTCGCATTTTCAAGAGAGATTAAGACGATTAATTAATTAATTAATTAATTAATTAAT | 1063 |
| Qy | 1109 | GGTGAACAGAGATGAGCAATATTTCAATGATGTTGTAACATGCTTTTACGCTTTTCGGTT | 1168 |
| Db | 1064 | GATGCAAGGAGAGAAATATATTTCTAGATGCTTCRACTTTTGCAATGSCCTTTTGAAT | 1123 |
| Qy | 1169 | ATTAAGATCAATGGGTATGAAGTTTCCCGATGCCATTTGGCTGAAATTAATA----- | 1222 |
| Db | 1124 | GTTACGTTTGAAGGATATGATGTTTCTTCAGACCAAGTTGACTCAATTTTCAGAAGATAT | 1183 |
| Qy | 1223 | -----TGAAATAGCTTTGAAGAGCAATATGACGCTCTTCAAAACATATCA | 1267 |
| Db | 1184 | CTTTCCCAATGGCTTTGGAGGATTAATTAAGACTTCGTTGCTCGCTGGAGTTATATA | 1243 |
| Qy | 1268 | TGCGTCACATATATTAATTAACAGAGGATTTATCTTCT---GGAAAAAATACTTGAAGTC | 1324 |
| Db | 1244 | GGCTCTCAGATTAATCACGCACCCCGATGAATCTGTTCTGMAAATATAAATCTTGGAC | 1303 |
| Qy | 1325 | AGCTGATTTCTCAAGAGATATATCCACTGATTCAAACAGGCTTTCTAAA----- | 1376 |
| Db | 1304 | TAGTCGTTTCTGAGCATGGATTAATCTAGTGATTTAGTTTGGTCTGATAGAACCGATAG | 1363 |
| Qy | 1377 | TTAATTCACAAAGAGGTGGAAAAATGCTCTTAAGTTCCCTATCAATAACGGTTTGAAGC | 1435 |
| Db | 1364 | TGTTCTTAAACBAGAGGCTGTTAATGCTCTTGAGTTCCCTATTAATCAACTCTAGAAGC | 1423 |
| Qy | 1436 | CATAAACCTAGACGAAATATACAGCTTTAATGATGACAAATACAGAAATTTCTGAAAC | 1495 |
| Db | 1424 | CCTAATAAGTAAGAGGCAATGGAAAGTTACAGTGGAGACATTTGTGAGGATTTCAAAATC | 1483 |
| Qy | 1496 | TACATATCACTCATCAAAATATAGTAACACATGATTACCTTAAGTTGGCTTGTGTAAGATTT | 1555 |
| Db | 1484 | GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTCTGGAACTTGTCTGTAGAGATTT | 1543 |
| Qy | 1556 | CTACACCTGCCAATCTATTTATCTGTGAAGAAATTAAGAGTCTTGAAGGTGGGTGTAGTA | 1615 |
| Db | 1544 | CAATACCTCGCAACGCAATTCATCTTAAGAACTCGAAGAGCTTCAAGATGGGTGTTGA | 1603 |
| Qy | 1616 | GAATAAGTTGGACAGCTCAAGTTGTTAGGCAAAAGACCGCTACTGTTATTTCTCTGT | 1675 |
| Db | 1604 | AAAAAATTTGGACGAGTTGAAATTTTTCAGACTGCACCTAGGTTACTGCTATTTTGTCTG | 1663 |
| Qy | 1676 | TGCTGCAACACTTCTGCTCCGCAATTAATCAATGCGCTATTTCAATGGGCCAAAAATCG | 1735 |
| Db | 1664 | GGCAGCGCCCTTACTGATCCCTGAATCTCATGATGCTCCATAGCTCGCATAGCTGGCACA | 1723 |
| Qy | 1736 | CATATTAACTACAGTAGTTGATGACTTTTGTGATATCGGTGGTACAAATCGATGAATGAC | 1795 |
| Db | 1724 | TGTGCTCAGACCGTGGTTGATGATTTCTATGATGGTGGAGGATCTGAAGAGGAAATGGA | 1783 |
| Qy | 1796 | CAACTGATTAATGTTTGAATAATGGAATGATGATGCTCGACAGGATTTGTTTTCAGA | 1855 |
| Db | 1784 | TAACTTTATAGAATTTGGTGGAAAAAGTGGATCTCGTGGGGAAGTGGTTTACTGTTTCCA | 1843 |
| Qy | 1856 | GCATGTTCCGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGGATTTGGAGATGAAGC | 1915 |
| Db | 1844 | GGACGTTGAGATTTGATTTCTTCTGACTGCACAGCAAGTTTGTGAATTAAGAGAGAGAGC | 1903 |
| Qy | 1916 | TTTTAAATGGCAAGCGCGATGTAACACTAGCCATGTTTATTCAAAACCTGTTGTTGAACTAAT | 1975 |
| Db | 1904 | TTTAGTATGCAAGGACGCGAGTGTATGAGGAATGTTATCGATGGTTGGTCTGCTGCT | 1963 |
| Qy | 1976 | GAATGATATTTGAGAGAGCTATATGACACAGAGATGCTTATGTCGCAACATTAATAATGA | 2035 |
| Db | 1964 | GAAGGTGATGAGAAAGAGAGCTGAATGGTTCGACAAATAAGGTAGTGCCTCAATGGGTGA | 2023 |
| Qy | 2036 | ATATATGAAAAACGCTTACGTGCTTTCATTTAGCCCGATTTGTCAGCCGCTATTTA | 2095 |
| Db | 2024 | ATATATGAAACAGCCCATGATCATTCGGTTGGACCTTAAATCTTCCATGCTCTT | 2083 |
| Qy | 2096 | CTTTGTGGGCCCCAAATTAACAGAGAGATTTGTGAAAGCTCTGAATATCATTAATCTATT | 2155 |
| Db | 2084 | CTTTGTGGACCTTAACTCTCAGAGGAATGATTGGAAGCTGTGAATACCAAGAGTTATA | 2143 |
| Qy | 2156 | TAAGCTAATGAGCACGCGAGCTGACTTCTTAACGATATCCATAGCTTCAAGAGGGAATTT | 2215 |
| Db | 2144 | TAAGCTGATGACACTGCTGCTCCCTTAAGATGATATTCGATCTTACGATGAGAGATG | 2203 |
| Qy | 2216 | TAAGGAAGCAAAATTAACCGCGTAGCATTTGCAATTTGAGTAACGGAAGAGTGGGAAAGT | 2275 |
| Db | 2204 | CAAGAGGGAAGCTGAATATTTCTGCTCTGTGGATGATTTGATGGCGGTGTTAATGTCAC | 2263 |
| Qy | 2276 | GGAAGAGAGGTTTGTGGAGAGATGATGATGATTAATAAAACAGAGGAAGAAATTAAT | 2335 |
| Db | 2264 | CAAGAGGAGGCCATTTGAAGCAATTAAGGGGATTTTGAAGGGCGATTAAGAGAGCTGCT | 2323 |
| Qy | 2336 | GAAATTAATTTTGAAGAAATGATGATGATTTCTCTAGAGCTTGTAAAGATGCATTTTG | 2395 |
| Db | 2324 | GGGTTAGTTTTCAGAGAGAACACTCAA---TTCAAGAGCTTTGAGGATTTGTTCTG | 2380 |
| Qy | 2396 | GAACATGCTGACGCTGTTGAAATTTTTCAGCAACAGATGACGGGTTTCTCGAAACAC | 2455 |
| Db | 2381 | GAAATGATGCTCATCTGATCTAATTTTACATGGAAGATGATGGGTACACTTCAAATAG | 2440 |

QY 2456 GATTCTGTGATCTGTAAGGACATCATTTTAAACCCGTTGGTCTTG 2502
 Db 2441 GTTGTGAACACTGTAAGGACCATGTTTGAACAACCCATGGATCTGG 2487

RESULT 2

US-08-727-308-3
 ; Sequence 3, Application US/08727308
 ; Patent No. 6020176
 ; GENERAL INFORMATION:
 ; APPLICANT: YUJI KAMIYA et al.
 ; TITLE OF INVENTION: KAURENE SYNTHASE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/727,308
 ; FILING DATE: October 8, 1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2658 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-727-308-3

Query Match 28.0%; Score 782.2; DB 3; Length 2658;
 Best Local Similarity 61.3%; Pred. No. 1.1e-191;
 Matches 1390; Conservative 0; Mismatches 823; Indels 54; Gaps 6;

QY 275 AACTAATCCCACTAATCTGATCATGTATACAAACAAAGACGATCCAAACAGTTTAA 334
 Db 236 AACTAAAACCTGGAGCTTTGCACTTTGAAGAAACAAAGAAAGAAATTAAGAAATTTGTCGA 295
 QY 335 AAATGTAGAATTTCTGTTTCTTCATATGACACAGCATGGTAGCCATGGTCCCTTCTCC 394
 Db 296 CAAGTTGAACCTTCAGTTTCTGATATGATGTCATGCGATGGTGGCAATGGTTCCTCTCC 355
 QY 395 AAACCTACCCAAATCGCCTTGTTCCTCCAGTGTCTCAATTTGGTTAATTAATATACAGT 454
 Db 356 AAACCTCTCAACCAACCTCTTTTCCCGAGTGTATAAAGTGGTATAGTAGTCAACA 415
 QY 455 TAATGATGGTTCATGGGTCTTGTTAATACACACTATAAATCATATCAACCGTTGCTTAA 514
 Db 416 TGCTGATGGCTCATGGGCGCTTACTCCCAAC-----GATCAGTTGCTGATGAA 463
 QY 515 AGATTCTCTATCTCAACATAGCATGTATGTTGCAATTAAGAGATGGATGTTGGGA 574
 Db 464 GGCCAAATCTCTTATCTACATAGCATGTGTTCTTACTCTTAAGCGGTGGAATATTGGCA 523

QY 575 AGATCAAAATAATAAAGGTCTAAGTTTATGAGTCBAATCTTGGCTTCTGAGTACTGAAAA 634
 Db 524 TGATCATATGACCAAGGCCCTTGATTTTATCAAGTCTAATATAGCTTCTAGCTACTGATGA 583
 QY 635 AAGTCAACCATCTCCCATTTGGTCTTGGACATCATATTTCTCGTTTGGTCTGAGTATGCGAA 694
 Db 584 GAAACCAAGCTTCTCGGTGGGATTTGACATTAATTTCCCTGGCATGATTTAGTAGTCTAA 643
 QY 695 AAACCTGGACATAAACCTCTCTTTCAAAAACAACAGATTTTGTGTTGATGCTACATAAGAG 754
 Db 644 AGACTTGAATTTGAATCTACCTTGGCACCGACGACGTCGATGCTTGGTTCGAAAGAA 703
 QY 755 GGAATTTGGAGCAAAAA-----AGATGCCATTCAAATGAGATGGATGATGATCTGCGTA 808
 Db 704 AGATTGGAGCTGAGAAAGCTGCAGAAAGCAACTCTGAAGTGGAAAGGCTTATTTAGCGTA 763
 QY 809 TATCTCTGAAGGACTCGGTAATTTATATGATTTGAATATGTTGAAGAAATATCATAGTAA 868
 Db 764 TGTTCAGAAGGAATTTGAAAGTTACAGACTGGGATATGTCATGCAATATCAAGAA 823
 QY 869 AAATGGTCTGTTTCAACTCAACATCAACAGCAAGCTCTGCTTTTATTAATCAATCAAAA 928
 Db 824 GAATGGATCACTGTTTAAATTTCCATCCACTACCGCAGCGGCTTTTATGCAATAGAAATGA 883
 QY 929 TCCTGGTTGCTTAATTAATTTAAATTTCACTTTTGGACAAGTTTGTATGATGCTCCCAAC 988
 Db 884 TGATGGCTGTTTGAATTAATCTTCGCTCACTCTTACAAAAGTTTGAATGCTCAGTTCCAC 943
 QY 989 AGTTTATCCTCATGATTTATTTATCCGACTTTCTATGTTTGACACAAATGAAAGATTAGG 1048
 Db 944 AATATATCCTCTTGATATATATGCTCGATTACATGTTGATAGCTTCAAAATTCGG 1003
 QY 1049 AATTTCACACATTTGAGAGTGGAAATTAATAATGTTTATAGTAAACATACAGATGTTG 1108
 Db 1004 AATTGCTGGCAATTTCAAGAGGAGATTTAGAAGCGTATTAGATGAAACTTACAGGTGTTG 1063
 QY 1109 GGTGAAACGAGATGAGCAAAATTAATCATGAGTGTGTAACATGCTGCTTTAGCTTTTGGTT 1168
 Db 1064 GATCGAAGGAGAGAAATATTTCTTAGATGCTTCAACTTGTGCAATGGCTTTTCSAAT 1123
 QY 1169 ATTAAGGATCAATGGGTATGAAGTTTCCAGATCCATTTGGCTGAAATTAATA----- 1222
 Db 1124 GTTACGTTTGAAGGATATGATGTTTCTTACAGCAGTGTGACTCAATTTTACAGAGATAT 1183
 QY 1223 -----TGAATAGCTTTGAAAGACGAATATGACAGCTCTTTGAAAACATATCA 1267
 Db 1184 CTTTCCCAATTTGGAGGATATTTAAAGACTTCGGTCCCTCGCTGGAGTTATATAA 1243
 QY 1268 TGCGTCAATATATATATACCAAGAGGATTTATCTTCT---GGAAACAAATCTTGAAGTC 1324
 Db 1244 GGCCTCTCAGATTAATACGCCCCCGATGAATCTGTTCTGGAAATATAAATCTCTTGAC 1303
 QY 1325 AGCTGATTTCTCAAAGAGATAATATATCCACTGATTTCAAACAGGCTTTCTAAA----- 1376
 Db 1304 TAGTCGTTTCTGAAGCATGGAATATCTAGTATGATTTGTTGTTGCTGATAGAACCGTAG 1363
 QY 1377 -TTAATTCACAAAGAGTGGAAATGCTCTTAAAGTTCCTTATCAATACCGGTTTGAACG 1435
 Db 1364 TGTGTTTAAACAGAGGCTGTTAATGCTCTTGTAGTTCCTTATATGCACTTAGAACG 1423
 QY 1436 CATAAACACATAGAGAAATATACAGCTTTTACAAATGTAGACATACAGAAATCTTGAAC 1495
 Db 1424 CCTAATAGTAAGGGCAATGGAAGTTACAGTGGAGACATTGTGAGGATTTCAAAATC 1483
 QY 1496 TACATATCACTCATCAATATTAAGTACACTGATTAAGTGGTGGCTGTTGAGGATTT 1555
 Db 1484 GCCATATGCTCTTAAATTTTGCCCATCAAGATTTTCTGGAATCTGCTGTAGAGGATTT 1543
 QY 1556 CTACACCTGCAATCTATTTTATCGTGAAGAAATTAAGAGTCTTGAAGAGTGGGTGTTAGA 1615
 Db 1544 CAATACCTGCAAGGCAATCTTCTTAAAGAACTTGAAGAGCTTCAAGAGATGGGTGTTGA 1603
 QY 1616 GAATAAGTGGACCGAGCTCAAGTTTGTAGGAAAGACCGCTTACTGTTATTTCTCTGT 1675

Db 1604 AAACAAATGGAGAGTGAATTTTTCAGACTGCACCTAGGTAATGCTAATTTTGCTGC 1663
QY 1676 TGCTGCAACATTTTCGTCCTCCGAAATATACAGATCGGCGTATTTTCATGGCCAAAATGG 1735
Db 1664 GCGAGCAACCTTACTGATCTCTGAACTTTCATGCTCGCATAGATGGGCAAAAATGG 1723
QY 1736 CATATTAACACTAGTAGTGAAGCTTTTTCATATCGGTGTACAAATCGATGAATGAC 1795
Db 1724 TGCTGCTCAGACCGTGTGTGATGATTTCTATGATGGTGAGGATCTGAAGAGAAATGGA 1783
QY 1796 CAACCTGATTTCAATGTGTGAAAAATGGAATAGATGCGAACAAGGATTTGTTGTCAGA 1855
Db 1784 TAACCTTATAGAAATGGTGAAAAAGTGGATCCTGATGGGAAGTGGGTACTGTTCCAA 1843
QY 1856 GAGTGTGGGATTTTATTTTATAGCATATAAAGATGCAATCTGTTGATTTGGAGATGAAGC 1915
Db 1844 GCGAGTGTGAGATTTGATTTCTTGCACATGACACAGATTTGTGAAATAGGAAGAGAGC 1903
QY 1916 TTTTAAATGGCAAGCGCGATGTAACTAGCATGTTTCAAACTTGGTTCGAACTAAT 1975
Db 1904 TTTAGTATGCAAGGACCGAGTGTATCGATGTTATGAGGATGTTATCGATGTTGGTCTGCT 1963
QY 1976 GAATAGTATGTTGAGAGAGCTATATGGAACAAGATGCTTATGTCGCCAACATTAAGA 2035
Db 1964 GAAGGTGATGAGAAAGGAAGCTGAATGGTTCGACAAATAAGGTAGTGCCATCAATGGGTGA 2023
QY 2036 ATATATGGAAGACGCTTACGTGTCAATTTGATTTAGCCCGATGTCACCGGCTATTTA 2095
Db 2024 ATATATGGAAGACGCTTATCAATCGGTGAGACCTTAATCTTTCCATGCTCTT 2083
QY 2096 CTTTGTGGGCCCCAATATATCAGAGGAGATTTGAAAGCTCTGAATATCAATCTATT 2155
Db 2084 CTTTGTGGACCTTAACTCTCAGAGGAATGATTTGGAAGCTGTGAATACCAAGTTATA 2143
QY 2156 TAAGCTTAATGAGCAGCAGCGGTGCACTTCTTAAACGATATCCATPAGTTTCAAGAGGAAT 2215
Db 2144 TAAGCTGATGAGCAGCTGCTGCTTAAAGATGATATTCGATCTTACGATAGAGAAATG 2203
QY 2216 TAAGAGGCAATTAACCGCGGTAGCATTTGATTAAGTACGGAAGAGTGGGAAAGT 2275
Db 2204 CAAAGAGGAAAGCTGAATATTTCTGCTCTGTGGATGATGATGGCGGTGTTAATGTCAC 2263
QY 2276 GGAAGAGAGGTGTTGAGAGATGATGATGATTAATAAACAAGAGGAAGAAATTAAT 2335
Db 2264 CAAAGAGAGGCAATTAAGCAATTAAGCGGATTTGAGAGCGGATAGAGAGCTGCT 2323
QY 2336 GAAATTAATTTTGAAGAAAATGGTAGCATTTGTTCTTAGAGCTTTGTAAGATGCAATTTG 2395
Db 2324 GCGGTTAGTTTTCAGAGGAACACTACAA---TTCCAAGAGCTTTGTAAGGATTTGTTCTG 2380
QY 2396 GAACATGTGTACGCTGTTGAATTTTAAAGCAAGATGACGGGTTTACTGGAACAC 2455
Db 2381 GAAATTAATGTCATTTGTAATCTATTTTACATGAGAGATGATGGGTACACTTCAATAG 2440
QY 2456 GATTTCTGTACTGTGAAGACATCATTTTCAACCCGTTGGTGTCTG 2502
Db 2441 GTTGAATGACACTGTAAAGCAATGTTTGAACCAATGATGCTGG 2487

RESULT 3

US-09-614-912-23
; Sequence 23, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude

; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: B81378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 2403
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1049)
US-09-614-912-23

Query Match 16.2%; Score 453.2; DB 4; Length 2403;
Best Local Similarity 54.9%; Pred. No. 6.9e+107;
Matches 1119; Conservative 0; Mismatches 808; Indels 111; Gaps 7;
QY 513 AAGAGATTCCTATCTTCAACATTAGCATGTTTGTGCAATTAATAAGATGGAATGTTGG 572
Db 72 AAGAGTTCCTATCTCCATCCAGTTGGCATGTTTGTGCGTTGAAGATGGAATGTTGC 131
QY 573 GAAGATCAATAAATAAAGTCTAAAGTTTATGAGTCAAAATCTTCTCCTCAGTACTGAA 632
Db 132 AGAGAGAACATTTGGAGAGGACTGCATTTTCATCGGAGGAATTTCTCTGTGCTATGGAC 191
QY 633 AAAAGTCAACATCTCCCATTTGTCATCATATTTTCCCTGTTGTTGTTGTTGTTGTTGTTG 692
Db 192 GAGCAGTTCACTTCTCTATAGTTTCAACTTCACTTCTCTGTTGTTGTTGTTGTTGTTGTTG 251
QY 693 AAAAAGTTCGACATAAACCTCTTCAAAAACAAACAGATTTTGTGTTGTTGTTGTTGTTGTTG 752
Db 252 ATTGATATGGTTTGAATTTCTGTGAACAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTG 311
QY 753 AGGAAATGGAGCAAAAAGATGCC-----ATTCAATGAGATGGAATGATGATGATGATGATG 803
Db 312 CGGGAGATGGAATTTGAAAGGGCTGGCTGTGGATAGTTTCTTTTGGAAAGAAAGCATATATG 371
QY 804 GCGTATATCTCTGAAGGACTCGTAAATTTATATGTTGGAATATGTTGGAATATGTTGGAATATCAG 863
Db 372 GCTTTTATCCAGAGGATTCGGAATATGCTGGAATGGAATGGAATGGAATGGAATGGAATG 431
QY 864 ATGAAAATGTTGTTCTGTTTCAACTCACTCAGCAACAGCTGCTGCTGTTTCAATTAATCAT 923
Db 432 AGGAAGATGGAATCATTTGTTGAGCACTCTTCCACATCTGCTGTTGTTGTTGTTGTTGTTG 491
QY 924 CAAAATCCTGTTGTTCTTAATTTAAATCTACTTTTGGACAAGTTTGGTAATGCAATC 983
Db 492 TACAACGACCAAGCCCTTCAATACCTAAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 551
QY 984 CCAAGATTTTATCTCATGTTTATTTATCCGACTTTTCTGTTGTTGTTGTTGTTGTTGTTGTTG 1043
Db 552 CCAGCAATGATCTTCAAGGGTCAATTTGTTGAGTTGTTGTTGTTGTTGTTGTTGTTGTTG 611
QY 1044 TTAGGAATTTTCAACACCATTTTCAAGATGGAATTAATAATGTTTATGATGAAACATACAGA 1103
Db 612 ATGGGAATTTTCTCAGCGCTTTTGTGAGTGAATAGAAAGCATCTCTGGACATGGCATCAAT 671

QY 1104 TGTGGGTGGAACGAGATGAGCAAAATATTCATGGATGTTGTAACATGTCGTTTACGCTTT 1163
Db 672 TCGTGTACAAATGATGAGGAATCTCATGATGAGCATAGCAACATTTGCAATGCAATTT 731
QY 1164 CGGTTATTAAGCATCAATGGGTATCAAGTTTCCAGATCCATTTGGCTGAAATTAATTAAT 1223
Db 732 CCGCTTTTGGAGATGAATGGTTTACGATGTTTCCCTCAGATGAGCTGTCTCACGTTGCTCGA 791
QY 1224 G-----AAATVAGCTTTTGAAGAGCAATATGTCAGCTCTTGAA 1259
Db 792 GCTTCCACTTTCATGATCTACTACAGGATATTTAAATGATACAAATCCCTACTGGAA 851
QY 1260 ACATATCATGCTGCA---TATATATACCAAGAGGATTTATCTTCTGGAAAAAATAATC 1316
Db 852 TTGTACAAGACCTCAAAAGTCACCTTATCAGAAAAAGATCTGATCTTAGATCGCATAGGT 911
QY 1317 TTGAAGTCAGCTGATTTCTCTCAAGAGATAATATCCAC----- 1354
Db 912 TCCTGGTCTGGCACTTATTAAGAGATAAGATGTGCTGTAGTAAAGGTGCAAAAGACTCGA 971
QY 1355 -----TGATTCAAACAGGCTTTCTAAATTAATTCACA----- 1386
Db 972 TTTTGGAGAGATGCTGCAACAAATTTAAATCTCAATTCACCTTGGAGGTTCACTT 1031
QY 1387 -----AAGAGTGGAAATGCTCTTAAGTTCCCTATCAATACCGGT 1427
Db 1032 TATGTCGTGTTTGTGNTTTCAGATCGAGTATGCTGTTAATTTCCCTTGTATTCACA 1091
QY 1428 TTAGAAGCATAAACACTAGACGAATATACGCTTTTACATGATGACAAATPACAGAAAT 1487
Db 1092 CTGGAGCTCTAGAACACAGAGAAACATCGAACATTTTGATGCTGGGGTCTCTGTGATG 1151
QY 1488 CTGAAATCTACATATCACTCATCAATATTAAGTAACTGATTAAGTAAAGTTGCTGCTT 1547
Db 1152 CT---AACAAATAATCTCATCTTTCTGATCAATCAAGAAATTCCTAGCTTTGCGAGTC 1208
QY 1548 GAAGATTTCTACCTGCAATCTATTTATCGTGAAGAAATTAAGGCTTTGAAAGGTGG 1607
Db 1209 GAAGATTTCACTTTCTCTCAACGCTTTTACCGGATGAATTCGCGATCTTGATAGTTGG 1268
QY 1608 GTGTAGAGATAGTTGGACCGCTCAAGTTTGTAGGCAAAAGACCGCTCTGTTAT 1667
Db 1269 GTGAAGAGAACAGCTGGACCGCTACAAATTTGCTGGCAAGAACTGACATATTTGCTAT 1328
QY 1668 TTCTCTGTGTGCAACACTTTGCTCTCCGAATATCAGATGCGGCTATTTTCATGCGCC 1727
Db 1329 CTGCTGCTGCTGCTACCGATTTTCTCTGAATTTGCTGAGCTGCGATTTTCATGCGCC 1388
QY 1728 AAAATGGCATATTAACATACAGTATGTTGATGACTTTTGTGATATCGGTGGTACAAATCGAT 1787
Db 1389 AAAATGGTGTCTCTCACAACTGTGTTGATGACTTCTTCGATGTTGGTGGATCAAAAGAA 1448
QY 1788 GAATGACCAACCTGATTCATGTTGTCAAAATGGAATGATGATGTCGACAAAGATTTGT 1847
Db 1449 GAATGAAACCTGATGATGACTAGTGTGAGAAATGGCATGGGCAACCATGCGATTTGAGTTC 1508
QY 1848 TGTTTCAAGCATGTTTGGATTTTATTTTATGCAATTAAGATGCAATCTGTTGGATTTGA 1907
Db 1509 TATTCGGAAACAGGTGAAATAGTATTTTCTGCTATTTATACACAGTGAACCATCTTTGA 1568
QY 1908 GATGAAGCTTTTAAATGGACCGCGGATGTAATAGCCATGTTATTCAAACTTGGTTG 1967
Db 1569 GCAATGGCTTCTGACGACACAGGCGGTGATCTTCAAAACCATCTAGTAGAAATATGCTG 1628
QY 1968 GAACATATGATAGTATGTTGAGAGAGCTATATGGAACAGATGCTTATGTCGCAACA 2027
Db 1629 GATTTGTTAGATCTATGATGTTGAGGAGATGAGGAGAGATGCCAATATATGACCAACA 1688
QY 2028 TTAATGAATATATGAAAAAGCTTTAGCTGTGCTTATTTGATTTAGGCGGATTTGCAAGCG 2087
Db 1689 GTTGAAGATACATGACAAATGCTGTTGCTCATTTGACATGGGCCCAATTTGCTCCCA 1748
QY 2088 GCTATTTACTTTTGTGGGCGCAAAATATACAGAGGAGATTTGTTGAAAGCTCTGAATATCAT 2147

RESULT 4

US-09-614-912-33

; Sequence 33, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

; APPLICANT: Lee, Jian Ming

; APPLICANT: Sakai, Hajime

; APPLICANT: Weng, Zude

; APPLICANT: Cai, Perry G

; APPLICANT: Anderson, Shawn

; TITLE OF INVENTION: Plant Metabolism Genes

; FILE REFERENCE: BB1378 US NA

; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: 60/143,401

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/143,412

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/146,650

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/170,906

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: 60/172,959

; PRIOR FILING DATE: 1999-12-21

; PRIOR APPLICATION NUMBER: 60/172,946

; PRIOR FILING DATE: 1999-12-21

; NUMBER OF SEQ ID NOS: 204

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 33

; LENGTH: 1157

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (40)

; NAME/KEY: unsure

; LOCATION: (44)

; NAME/KEY: unsure

; LOCATION: (60)

; NAME/KEY: unsure

; LOCATION: (64)

; NAME/KEY: unsure

; LOCATION: (712)

; NAME/KEY: unsure

Db 1749 GCATTTGATTTTGTAGGCAAGAGCTATTAGAGCATGCTGTCAAAGATGAAGAGTACGAT 1808
QY 2148 AATCTATTTAAGCTAATAGACGACGCGGTGAGCTTTCTAAACGATPATCCATAGCTTCAAG 2207
Db 1809 AAAATTTATTTAGCTAGTGAGCACTTGGCGGAGGCTCTCAATGACTACCAAAAGTTTAGAG 1868
QY 2208 AGGGAATTTAAGGAGGCAATTAACCGGCTAGCAATTTGCAATTTGAGTAAACGAGAAAGT 2267
Db 1869 AGGGAAGGCAACCGAGGGAAGCTGAATAGTGTCTTCTACTGTGCTCCACAGTGGTGT 1928
QY 2268 GGGAAAGTGGAAAGAGGTTCTGGAGGAGATGATGATGATGATGATGATGATGATGATGATGAT 2327
Db 1929 TCTATGTCCTAAGAGCGCTAAAGAGGCAATGAGAGTCCATAGAGCTGTCTAGGAGA 1988
QY 2328 GAATTAATGAATTAATTTTGAAGAAATGGTGTAGCATGTTCTTAGAGCTTTGTAAGAT 2387
Db 1989 GACTTGTCTAAGATTTGGTTCT---CAGGAAAGAAAGTGTGTTCTTAGCCATGCAAGGAG 2045
QY 2388 GCATTTGGAAACATGTGTCAGCTGTGTAATTTTCTTACGCAACGATGACGGGTTTA 2445
Db 2046 CTCCTTGTGAAGATGTGTAAGATCTTCCACTGTTTACTCTCAGATGATGGATTTA 2103

| | | |
|---|-----------|------------------|
| ? | LOCATION: | (721) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (723) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (732) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (738) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (757) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (763) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (782) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (809) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (813) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (824) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (826) .. (827) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (849) .. (850) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (858) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (863) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (872) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (874) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (888) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (930) .. (931) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (938) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (942) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (961) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (967) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (971) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (981) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (988) .. (989) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (993) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (999) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1024) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1028) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1031) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1035) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1040) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1049) .. (1050) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1068) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1089) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1105) .. (1106) |
| ? | NAME/KEY: | unsure |
| ? | LOCATION: | (1121) |

; APPLICANT: No. 6468721, Joseph P.
 ; APPLICANT: Starks, Courtney M.
 ; APPLICANT: Manna, Kathleen R.
 ; TITLE OF INVENTION: SYNTHASES
 ; FILE REFERENCE: 07678-025001
 ; CURRENT APPLICATION NUMBER: US/09/398,395A
 ; PRIOR FILING DATE: 1999-09-17
 ; PRIOR APPLICATION NUMBER: 60/100,993
 ; PRIOR FILING DATE: 1998-09-18
 ; PRIOR APPLICATION NUMBER: 60/130,628
 ; PRIOR FILING DATE: 1998-04-22
 ; PRIOR APPLICATION NUMBER: 60/150,262
 ; PRIOR FILING DATE: 1999-08-23
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 55
 ; LENGTH: 2861
 ; TYPE: DNA
 ; ORGANISM: Abies grandis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (3)...(2606)
 ; OTHER INFORMATION: abietadiene synthase
 ; US-09-398-395A-55

Query Match 7.0%; Score 194.8; DB 4; Length 2861;
 Best Local Similarity 47.8%; Pred. No. 3.1e-40;
 Matches 866; Conservative 0; Mismatches 85; Indels 87; Gaps 7;

| | | | |
|----|-----|---|------|
| QY | 337 | ATGTAGAAATTCCTGTTCTTCATATGACACAGCATGGGTAGCCATGTCCTCCAA | 396 |
| DB | 397 | ATGGGAAACGAAATCCCTCTGCATATGACACTGCTTGGGTAGCAAGGATTCAGCAGTTG | 456 |
| QY | 397 | ACTCACCCAAATCGCTTGTTCCTGAGTCTCAATGTTGTAATTAATCAATCAGCTTA | 456 |
| DB | 457 | ATGGCTCTGACAAACCTTCACCTTCTGAGAGCGTTGGAATGATTTCCAAATCAGTTGA | 516 |
| QY | 457 | ATGATGGTTCATGGGCTTGTGTAATCACACTCATATCAATCACCCGTTGCTTAAAG | 516 |
| DB | 517 | AAGATGGCTCTGGGGTGAAGGATTCCTACTTC-----TTGGCATAATG | 558 |
| QY | 517 | ATTCCTATCTTCAACATAGCATGATGTTGCAATTAAGAGATGGAATGTTGGGAG | 576 |
| DB | 559 | ACGAATCTGGCTACACTGCTGATGATATATACCTTTACCTCTGGGCTACTGGGAGA | 618 |
| QY | 577 | ATCAATAAATAAAGGCTTAAGTTTATTTAGTCAATCTTCTTCAGCTACTGAATAA | 636 |
| DB | 619 | CACAAGTACAGAAGGATTAATCTTCAGACACACAGCTGGAAAGATGAAG | 678 |
| QY | 637 | GTCACCAATCT---CCCATGGTTTGAATCATATTTCTGTTGCTTCAATGATGCGA | 693 |
| DB | 679 | CTGATAGTCATAGGCCAAGTGGATTTGAAATAGTATTTCTGCAATGCTAAAGGACTA | 738 |
| QY | 694 | AAACTTTGGACATAAACTCTTCAAAACAAACAGATTTAGTTGATGCTACATAGA | 753 |
| DB | 739 | AAATCTTAGGCTGGATCTGCTTACGATTTGGCAATTTGCCATTTCTGAAACAAATCATCGAAAGC | 798 |
| QY | 754 | GGGAATTTGGAGCAAAAAGATGCCATTCAAATGAGATGGATG-----GATACTGG | 804 |
| DB | 799 | GGGAGGCTAGCTTAAAGGATTTCCACTGATGTTCTATGCTTCCCAACACGTTAT | 858 |
| QY | 805 | CGTATATCTCTGAAGGCTCGGTAAATTTATATGATGGAATATGTTGGAAGAAATATCAGA | 864 |
| DB | 859 | TGTATTTCTTTGGAAGTTTCAAGAAATAGTAGACTGGCAGAAATATGAACTTCAAT | 918 |
| QY | 865 | TGAATAATGTTCTGTTTCACTCACATCAGCAACAGCTGCTGCTTCAATTAATCATC | 924 |
| DB | 919 | CCAGGATGATCATATTTCTAGCTCTCCGGCATCTACAGCGGCTGTATTATGCGGTACAG | 978 |
| QY | 925 | AAATCTGGTGTCTTAATTTAATTCACCTTTTGGACAGTTTGGTAAATGCACTCC | 984 |
| DB | 979 | GGAAACAAAAGTGTGGAATTTTGAACTTTGTTGTTGGAAGAAATTCGGAACCATGTGC | 1038 |

| | | | |
|----|------|--|------|
| QY | 985 | CAACAGTTTATCCTCATGATTTATTTATCCGACTTTCTATGTTGACACAAATGGAAGAT | 1044 |
| DB | 1039 | CTTGTCATATCCGCTTGATCTATTTGAACTGTTGTGGCGGTTGATACAGTTGAGCGGC | 1098 |
| QY | 1045 | TAGGAAATTTCAACACATTTACAGATGGAAATTTAAATAATGTTTATAGATGAAACATACAGAT | 1104 |
| DB | 1099 | TAGGTATCGATCGTCAATTTCAAAGAGGAGATCAAGGAAGCATTTGATTTATGTTTACAGCC | 1158 |
| QY | 1105 | GTGGGTGGAACGAG-----ATGAGCAAAATATTCATGGATGTTGTAACAT | 1149 |
| DB | 1159 | ATTGGACGAAAGAGGCATTTGGATGGCGAGAGAGAAATCCTGTTCTCTGATATGATGATA | 1218 |
| QY | 1150 | GTGCTTTAGCCTTTCGGTTATTAAGGATCAATGGGTATGAACTTTCCCGGATTCATTTGG | 1209 |
| DB | 1219 | CAGCATGGCTTTCGAATCTTGAGATTACATGGGATCAATGATATCTCCAGATGTTTAA | 1278 |
| QY | 1210 | CTGAAATTAATAAGAAATAGCTTTGAAAGACGAATATGCAGCTCTTGAACATATCATG | 1269 |
| DB | 1279 | AAACATTTAGAGATGAGAAATGGGAGTTCTTTTGGCTCTTGGGTCAAAACAGAGGAG | 1338 |
| QY | 1270 | CGTCACATATATTAAC-----AAGAGATTTATCTTCTGGAACAAATCT | 1317 |
| DB | 1339 | TTACAGACATGTTAAACGTCATCGTTGTTTACATGTTTCTATTTCCGGGAGAAACGATCA | 1398 |
| QY | 1318 | TGAAGTCAGCTGATTTCT-----CTCAAGAGATAATAT | 1350 |
| DB | 1399 | TGGAGAGCAAACTCTGTACCGAAAGGATATGAGGAATGCTCTGGAATTTGGAGT | 1458 |
| QY | 1351 | CCACTGATTTCAACAGCGCTTTCTAAATTTAAATTCACAAAGAGGTGGAATGCTCTTAAAT | 1410 |
| DB | 1459 | CCTTTGACAAATGGCTTTTAAAGAAATATTCGGGAGAGGTAGAGTATGCATCAAT | 1518 |
| QY | 1411 | TCCCTATCAATACCGTTTGAAGACGATTAACACATAGACGAAATATACAGCTTTACAAG | 1470 |
| DB | 1519 | ATCCCTGGCATAAGAGTATGCCAAGTTGGAGGTAGAACTATATTTGAAACTATGGC | 1578 |
| QY | 1471 | TAGACATACAAAGAAATCTGAAACTACATCTCATCAATTAATTTAGTAACTGAT | 1530 |
| DB | 1579 | CAGATGATGCTGCTTGGAAAACTGTATATATGATGCCATACATTTCGAATGAAAGT | 1638 |
| QY | 1531 | ACCTAAGTTTGGCTGTTGAAGATTTTACACCTGCCAATCTATTTATCGTGAAGAAATAA | 1590 |
| DB | 1639 | ATTTAGAACTAGCGAACTGACCTTCAATAGGTGCACTATACACCAACAGAGCTTC | 1698 |
| QY | 1591 | AAGCTCTGAAAGTGGTGGTAGAATAAGTTGGACCGCTCAAGTTTCTAGGCAAA | 1650 |
| DB | 1699 | AAGATCTTGAAGGTGGTGAATCATCCGGTTTTCAGCGATCTGAATTTCACTCGTACG | 1758 |
| QY | 1651 | AGACCGCTACTGTTATTTCTGCTGCAACACTTTGCTCTCCCGAATTTACAGATG | 1710 |
| DB | 1759 | GTGTGACGGAATAATATTTCTCCCGCATCTTTATCTTTGAGCCCGAGTTTCTTAAGT | 1818 |
| QY | 1711 | CGGTATTTATGGGCCCCAAATATGACATATTAACACAGTAGTTGATGATCTTTTTCGATA | 1770 |
| DB | 1819 | GCAGAGAGTTTATACAAAACTTCCAAATTTCACTGTTATTTTAGATGATCTTTATGACG | 1878 |
| QY | 1771 | TCGGTGTCAATTCGATGATTCACCAACTGATCTCAATGTTGTTGAAAAATGGAATGTAG | 1830 |
| DB | 1879 | CCCATGGATCTTTAGACGATCTTAAGTTGTTTACAGAAATCAGTCAAAAGATGGATCTAT | 1938 |
| QY | 1831 | ATGTCGACAAAGGATTTGTTTCAGAGCATCTTCGGAATTTATTTTATAGCATTTAAAGATG | 1890 |
| DB | 1939 | CACATAG---TGGACCAATCCCAACAAATGAAATATGTTTGTGGGTTTCTACAAATA | 1995 |
| QY | 1891 | CAATCTGTTGGATGGAGATGAAGCTTTTAAATGGCAAGCGCGCATGTAACTAGCCATG | 1950 |
| DB | 1996 | CTTTTAATGATATAGCAAAAGAGGAGCTGAGAGGCAAGCGCGCATGTCTAGGCTACA | 2055 |
| QY | 1951 | TTATTCAACTTGGTGGAACTAATGAATAGTATGTTGAGAGAGCTTATATGACAGAG | 2010 |
| DB | 2056 | TTCAAAATGTTGGAAAGTCCAACTTTGAAGCTTTACAGAAAGAGAGAGATGTTCTGAAG | 2115 |
| QY | 2011 | ATGCTTATGTCACATTAATGAATATATGAAACCGCTTACGTTGCTCATTTGCTATAG | 2070 |

Db 2116 CTAATAATGTGCGCATCTTCAATGAATACATAGAGATGCGAGTGTGTCATAGCAATTGG 2175
 Qy 2071 GCCCGATTGT 2080
 Db 2176 GAACAGTCGT 2185

RESULT 6
 US-09-887-586A-55
 ; Sequence 55, Application US/09887586A
 ; Patent No. 6495354
 ; GENERAL INFORMATION:
 ; APPLICANT: Chappell, Joseph
 ; APPLICANT: No. 64953541, Joseph P.
 ; APPLICANT: Starks, Courtney M.
 ; APPLICANT: Manna, Kathleen R.
 ; TITLE OF INVENTION: SYNTHASES
 ; FILE REFERENCE: 07678-025001
 ; CURRENT APPLICATION NUMBER: US/09/887,586A
 ; CURRENT FILING DATE: 2001-06-22
 ; PRIOR APPLICATION NUMBER: 09/398,395
 ; PRIOR FILING DATE: 1999-09-17
 ; PRIOR APPLICATION NUMBER: 60/130,628
 ; PRIOR FILING DATE: 1999-04-22
 ; PRIOR APPLICATION NUMBER: 60/150,262
 ; PRIOR FILING DATE: 1999-08-23
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 55
 ; LENGTH: 2861
 ; TYPE: DNA
 ; ORGANISM: Abies grandis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (3)...(2606)
 ; OTHER INFORMATION: abietadiene synthase
 US-09-887-586A-55

Query Match 7.0%; Score 194.8; DB 4; Length 2861;
 Best Local Similarity 47.8%; Pred. No. 3.1e-40;
 Matches 866; Conservative 0; Mismatches 857; Indels 87; Gaps 7;

Qy 337 ATCTAGAAATTTCTGTTTCTTCATATGACACAGATGGGTAGCCATGTCCTTCTCCAA 396
 Db 397 ATGGCGAAGCAATCCCTCTGCATATGACACTCTTGGGTAGCAAGATTCAGCAGTTG 456
 Qy 397 ACTCACCCAAATCGCCTTGTTCCTCGAGTGTCTCAATTGGTTTAAATTAATCACTTA 456
 Db 457 ATGGCTCTGACACACCTCACTTTCCTGAGACGGTTGAATGGAATTCCTCAAAATCAGTTGA 516
 Qy 457 ATGATGTTTCATGGGTCCTGTTGTAATCACTCATATCAATATCAATCCCGTTCCTTAAG 516
 Db 517 AAGATGGGTCTTGGGGTGAAGATTCCTACTTC-----TTGGCATATG 558
 Qy 517 ATTCTCTATCTCAATAGCATGATTCCTGCAATTAAGATGAAATGTTGGGGAAG 576
 Db 559 ACAGATATCTGGCTACACTTGCAATGATTAATACCTTACCCTCTGCGGTACTGGGGAGA 618
 Qy 577 ATCAAAATAAAGGCTAAGTTTATTCAGTCAATCTTGTCTTCACTTCACTTCAAAAAA 636
 Db 619 CACAAGTACAGAAAGGTATTGAATTCCTCAGGACACAAGCTGGAAGATGGAAGTGAAG 678
 Qy 637 GTCAACCATCT---CCCATGGTTTTCACATCATATTTCTCTGTTTCTGAGTATGCGA 693
 Db 679 CTGATAGTCATAGGCCAAGTGGATTTGAAATAGTATTTCTGCAATGCTAAAGGAAGCTA 738
 Qy 694 AAAAATGGGACATAAACTCTCTTCAAAACAAACAGATTTTAGTTTGTATGCTACATAAGA 753
 Db 739 AAATCTTAGGCTTGATCTGCTTACGATTTGCCATTCCTGGAACAATCATCGAAAAGC 798
 Qy 754 GGGAAATGGAGCAAAAAGATGCGCAATTCAAATGAGATGGATG-----GATACCTTGG 804

Db 799 GGGAGGCTAAGCTTAAAAAGGATTCCTCACTGATGTTCTCTATGCGCCTTCCAAACAGCTTAT 858
 Qy 805 CGTATATCTCTGAGAGACTCGGTAAATTTATATGATGGAATATGATGGAAGAAATATCAGA 864
 Db 859 TGTATTTCTTGGAGGTTTACAGAAATAGTAGACTGGCAGAAATATATGAACCTTCAAT 918
 Qy 865 TGAAAAATGTTTCTGTTTCACTCAGCATCAGCAACAGCTGCTGCTTTCATTAATCATC 924
 Db 919 CCAAGGATGGATCATTTCTCAGCTCTCCGGCATCTACAGCGGCTGTATTCATGCGCTACAG 978
 Qy 925 AAAATCCTGGTTGCTTAAATTTAAATTTCACTTTTGGACAAGTTTGGTATGCACTCC 984
 Db 979 GGAACAAAAAGTGTGTTGATTTCTTGAACCTTTGCTTTGAAGAAATTCGGAACCAATGTC 1038
 Qy 985 CAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGCTTGACACAAATGAAAGAT 1044
 Db 1039 CTTGTCATCTCCGTTGATCTATTTGAACGTTTGGCGGTTGATACAGTTGAGCGGC 1098
 Qy 1045 TAGGAAATTTACACCATTTTCAAGTGGAAATTTAAATTTTATAGTAAACATACATCAGAT 1104
 Db 1099 TAGGTATCGATCGTCAATTTCAAAGAGGAGATCAAGGAACATTTGGATTTATGTTTACAGCC 1158
 Qy 1105 GTTGGGTGAACGAG-----ATGACCAATATTTTCATGGATTTGTAACAT 1149
 Db 1159 ATTGGGACGAAGAGGCAATTGATGGCGAGAGAGAAATCTCTGTTCTGATATATGATATA 1218
 Qy 1150 GTGCTTTAGCCTTTTCGTTTATTAAGGATCAATAGGTTATGAAGTTTCCCGAGATCCATTTGG 1209
 Db 1219 CAGCCATGGGCTTTGAACTTTGAGATTTACATGATACATGATCTCAGATGTTTTAA 1278
 Qy 1210 CTGAAATTTAATGAATTTAGCTTTGAAAGAGCAATATGACGCTCTTGAACATATCATG 1269
 Db 1279 AAACATTTAGAGATGAGAATGGGAGTTCTTTTGTCTTTGGGTCAAACACAGAGAGGAG 1338
 Qy 1270 CGTCAATATATATAC-----AAGAGGATTTATCTTCTCGAAACAAATCT 1317
 Db 1339 TTACAGACATGTTAAACGCAATCGTTGTTGTTCATGTTTTCATTTCCGGGAGAAACATCA 1398
 Qy 1318 TGAAGTCAGCTGATTTTC-----CTCAAAGAGATTAATAT 1350
 Db 1399 TGGAGAAGCAAACTCTCTGTAACGAAGGTATCTGAGGAATGCTCTGGAATATGCGATG 1458
 Qy 1351 CCACATGATTTCAACAGGCTTTCTAAATTAATTTCAAAAGAGTGGAAATGCTCTTAAGT 1410
 Db 1459 CTTTTGACAAATGGGCTTTTAAAGAGATATTTGGGGAGAGGTAGATGCACTCAAAAT 1518
 Qy 1411 TCCTATCATACCGGTTTGAACGCAATAACACATAGACGAATATACAGCTTTTCAATG 1470
 Db 1519 ATCCCTGGCATAGAGATTCGCCAAGGTTGGAGGCTAGAAGCTATATTTGAAACTATGGGC 1578
 Qy 1471 TAGACAATACAGAAATTTCTGAAACATACATATCACTCATCAATATTTAGTAAACATGAT 1530
 Db 1579 CAGATGATGTGGCTTTGGAAACCTGATATATATGATGCCATACATTTTCAATGAAAGT 1638
 Qy 1531 ACCTAAGTTGGCTGTTTGAAGATTTCTACCTGCCAATCTATTTATTCGTGAGAGATTA 1590
 Db 1639 ATTTAGAACTAGCGAACTGGACTTCAATAAGGTGCAGTCTATACCAACAGAGAGCTTC 1698
 Qy 1591 AAGGCTCTGAAAGGTGGGTAGAGAATAAGTTGGACCAAGCTCAAGTTTGTCTAGGCAAA 1650
 Db 1699 AAGATCTCGAAGGTGGTGAATCATCCGGTTTCAGGATCTGAAATTTCACTCTGAGC 1758
 Qy 1651 AGACCCGCTACTGTTATTTCTGTTGCTGCAACACTTTCGTTCCCGAATTTACAGATG 1710
 Db 1759 GTGTGACGGAATATATTTCTCAGCGCATCTTTATCTTTGAGCCGAGTCTTTTAAAGT 1818
 Qy 1711 CGGATTTTCATGGCCAAATATGCGATATTAATACAGTATGATGATGATCTTTTTCGATA 1770
 Db 1819 GCAGAGAGGTTTATACAAAACCTTCCAAATTTCACTGTTATTTTAGATGATCTTTATGAGC 1878
 Qy 1771 TCGGTGTTCAATCGATGAAATTCACCAACCTGATTCATGTTTGAATTTGAAAAATGGAATG 1830
 Db 1879 CCCATGATCTTTAGACGATCTTAAAGTTGTTTCACAGATCAGTCAAAAGATGGGATCTAT 1938


```
QY 1591 AAGGCTTTGAAAGGTGGGTGTAGAGATAAGTTGGACCAAGTCAAGTTTGTAGGCAA 1650
Db 1699 AAGATCTTCGAAGGTGGTGGAAATCATCCGGTTTCACGGATCTGAATTTCACTCGTGAGC 1758
QY 1651 AGACCGCTACTGTTATTTCTGTTGTTGTTGCAACACTTTTCGTCCTCCCGAATTTATCAGATG 1710
Db 1759 GTGTACCGGAATATATTTCTCACCGGCATCTTTATCTTTGAGCCCGAGTTTCTAAGT 1818
QY 1711 CGCGTATTTCAATGGGCCAAAANGGCAATATTAACATACAGTAGTGTATGACATTTTGTGATA 1770
Db 1819 GCAGAGAGGTTTATACAAAACCTTCCAAATTTCACTGTTTATTTTATAGATGATCTTTATGACG 1878
QY 1771 TCGGTGGTACATCATGATGAATTCACCAACCTGATTCATGCTTGAAATAAGGATGTAG 1830
Db 1879 CCCATGATCTTTAGACGATCTTAAGTTGTTTACAGAAATCAGTCAAAAGATGGGATCTAT 1938
QY 1831 ATGTCGACAAAGATTTGTTGTTTCAGAGCATGTTCCGATTTTATTTTATAGCATTTAAAGATG 1890
Db 1939 CACTAG---TGACCAATGCCACAACAAATGAAATATGTTTGTGGTTTCTACAATA 1995
QY 1891 CAATCTGTTGGATGGAGATGAAGCTTTTAAATGGCAAGCGCGCATGTAACTAGCCATG 1950
Db 1996 CTTTAAATGATATAGCAAAAGAGGACGTGAGAGGCAAGGCGCATGTGTAGGCTACA 2055
QY 1951 TTAATCAAACTGGTTGGAATCAATCAATAGTAGTGTGAGAGAAGCTATATGCAAGAG 2010
Db 2056 TTCAAAATGTTGGAAAGTCCAACTGGAAGCTTACAGAAAGAGCAAGATGCTGTGAAG 2115
QY 2011 ATGCTTATGTGCAACATTAATGAATATATGAAACCGCTTTACGTGTCAATTTGCATTTAG 2070
Db 2116 CTAATATGTGCACTTCATGATATACATAGAGATGCGAGTGTGTCAATAGCATTTGG 2175
QY 2071 GCCCGATTGT 2080
Db 2176 GAACAGTGT 2185
```

RESULT 8

```
US-09-903-012B-55
; Sequence 55, Application US/09903012B
; Patent No. 6569656
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65696561, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/903,012B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 55
; LENGTH: 2861
; TYPE: DNA
; ORGANISM: Abies grandis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(2606)
; OTHER INFORMATION: abietadiene synthase
US-09-903-012B-55
```

Query Match 7.0%; Score 194.8; DB 4; Length 2861;
Best local similarity 47.8%; Pred. No. 3.1e-40;
Matches 866; Conservative 0; Mismatches 857; Indels 87; Gaps 7;

```
QY 337 ATGTAGAAATTTCTGTTTCTTATATGACAGACATGGGTAGCCATGTCCTTCTTCCAA 396
Db 397 ATGGGAAACGAATCCCTCTGATATGACATGCTGTTGGGTAGCAAGGATTCAGCAGTTG 456
QY 397 ACTCACCCAAATCGCTTGTTCCTGAGTGTCTCAATGGTAAATTAATCAAGCTTA 456
Db 457 ATGGCTCTGACAAACCCCTCACTTTCCTGAGACGGTTGAATGGATTTCTTCAAAATCAGTTGA 516
QY 457 ATGATGGTTTCATGGGTTCTTGTAAATCACACTCATATAATCACCCGTTGCTTAAAG 516
Db 517 AAGATGGGTTCTGGGTGAAGATTTCTACTTC-----TTGGCATATG 558
QY 517 ATTTCTTATCTTCAACATTTAGCATGTATGTTGCAATTAAGAGATGGAATGTTGGGAG 576
Db 559 ACAGAATACTGGCTACACTTGCATGTATTAATACCTTTACCTCTGGGCTACTCGGGAGA 618
QY 577 ATCAATTAATTAAGGCTTAAGTTTATGAGTCAAACTTCTGCTTCAGCTACTGAAAAA 636
Db 619 CACAAGTACAGAAAGGTATTGAATTTCTCAGGACACAAGCTGGAAGATGGAAGATGAAG 678
QY 637 GTCAACCATCT---CCCATTTGGTTTGGATCATCATATTTCTTGGTTTGGTGTGATGCGA 693
Db 679 CTGATAGTCAATAGGCCAAGTGGATTTGAAATAGTATTTCTTGCANTGCTAAAGAGCTTA 738
QY 694 AAAAATTTGGACATAAAACCTCTTTCAAACAAACAGATTTTATGTTTGTGCTTACATAAGA 753
Db 739 AAATCTTAGGCTTGGATCTGCTTACGATTTGCCATTTCTGAAACAAATCATCGAAAAAGC 798
QY 754 GGGAAATTTGGACAAAAAGATGCCATTCAAATGAGATGGATG-----CATACTTGG 804
Db 799 GGGAGGCTTAAGCTTTAAAGGATTCCTCAGTGTCTCTATGCCCTTCCAAACAGTTAT 858
QY 805 CQTATATCTCTCAAGGACTCGGTAATTTATATGATTTGGAATATGTTGGAAGAAATATCAGA 864
Db 859 TGTATTTCTTGGAGTTTACAGAAATAGTAGACTGGCAGAAATATATGAATCTCAAT 918
QY 865 TGAAAAATGGTTCTGTTTTCAACTCAACATCAGCAACAGCTGCTGCTTTTCAATTAATCATC 924
Db 919 CCAAGATGATGATCAATTTCTCAGCTCTCCGGCATCTACAGCGCTGTATTCATGCGGTACAG 978
QY 925 AAAATCCTGGTTGCTTAAATTTAAATTCATTTTGGACAAAGTTTGGTAAATGCAAGTCC 984
Db 979 GGAACAAAAAAGTGTGGATTTCTTGAACCTTTGCTTTGAAAGAAATTCGGAACCATGTGC 1038
QY 985 CAACAGTTTATCTCATGATTTTATTTATCCGACTTTTCTATGTTGTGACAAATTTGAAGAT 1044
Db 1039 CTTGTCACTATCGCTTGTATCTATTTGAACGTTTGTGGCGGTTGATACAGTTGAGCGGC 1098
QY 1045 TAGGAATTTTCACACCATTTTCAGAGTGGAAATTAATAATGTTTATGATGAACAATACAGAT 1104
Db 1099 TAGGTATCGATCGTCAATTTCAAAGAGGAGATCAAGGAAGCATTTGGATTTATGTTTACAGCC 1158
QY 1105 GTTGGGTGAAACGAG-----ATGACCAATATTTTCATGGATTTGTTAAACAT 1149
Db 1159 ATTGGGACGAAAGAGGCAATTTGGATGGGCGAGAGAGAAATCCTGTCTCTGATTTATGATGA 1218
QY 1150 GTGCTTTAGCTTTTCGGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATFCCATTTGG 1209
Db 1219 CAGCCATGGGCTTCCGATCTTGAATTTAGATTTACATGATATCTCTCAGATGTTTTAA 1278
QY 1210 CTGAAATTAATAATGAATTTAGCTTTGAAAGAGCAATATGACAGCTCTTTGAAAATATCATG 1269
Db 1279 AAACATTTAGAGATGAGAAATGGGAGTTCTTTTGTCTTTTGGTCAACACAGAGAGGAG 1338
QY 1270 CGTCAATATATTTATACC-----AAGAGGATTTTATCTTCTCGAAAAACAATCT 1317
Db 1339 TTACAGACATGTTTAAACGTCATCTGTTTGTTCATGTTTTCATTTCCGGGAGAAACGATCA 1398
QY 1318 TGAAGTCAGCTGATTTTC-----CTCAAAGAGATTAATAT 1350
Db 1399 TGGAAGAGCAAACTCTGTACGAAAGGTATCTGAGGAATGCTCTGGAATAATGTGGATG 1458
QY 1351 CCACTGATTCAAACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAAAAATGCTCTTAAGT 1410
```


1210 CTGAATTAATAAGTCTTGAAGACGAAATATGCAGCTCTTGAACATATCATG 1269
1279 AAACATTTAGAGATGAGATGGGAGTCTTTTGTCTTCTGGGTCAAAACACAGAGAGAG 1338
1270 CGTCACATATATATACC-----AAGAGGATTAATCTCTGGAACAACTATCT 1317
1339 TTACAGACATGTTAAACGTCATCTGTTTTCATGTTTTCATGTTTCGGGAGAAACGAICA 1398
1318 TGAAGTCAGCTGATTTTC-----CTCAAGAGATAATAT 1350
1399 TGAAGAGACAAATCTCTGACGAAAGGTATCTGAGGAATGCTCTGGAAAATGTGGATG 1458
1351 CCATGATTCACACAGGCTTCTAAATTAATTCACAAAGAGGTGGAATGCTCTTAAGT 1410
1459 CTTTTCACAAATGGGCTTTTAAAGAGATATTTTCGGGGAGGTAGAGATGACATCAAT 1518
1411 TCCCTATCAATACCGGTTTGAACGCAATAACACTAGACGAATATATACAGCTTTTCAATG 1470
1519 ATCCCTGGCATAGAGATGCCAGGTTGGAGGCTAGAGCTATATTTGAACACTATGGGC 1578
1471 TAGACAAATCAAGAAATCTGAAAACCTACATATCATCTCAATCAATATATAGTAACACTGAT 1530
1579 CAGATGATGTCGGCTTGGAAAACCTGTATATATGATGATGATGATGATGATGATGATGAT 1638
1531 ACCTAAGGTTGGCTGTTGAAGATTTCTACACCTGCGCAATCTATTTATCGTGAAGATTA 1590
1639 ATTAGAACTAGCGAACTGGGACTTCAATAGGTGCGATCTATACACAAACAGAGCTTC 1698
1591 AAGGCTTTGAAAGGTGGGTAGAGATAAGTTGACAGCTCAAGTTTGTCTAGGCAAA 1650
1699 AAGACTTTCGAGGTTGTTGGAATCATCCGTTTTCAGGATCTGAAATTTCACTCGTGAGC 1758
1651 AGACGCTTACTGTTTCTGTTGCTGCGTGCACACTTTTCTGCTCCGAAATATCAGATG 1710
1759 GTGTGACGGAATATATTTCTCACCGGCATCTTTATCTTTGAGCCGAGTTTCTTAAGT 1818
1711 CGGTTATTTTCATGGGCAAAATGGCATATTAACACTAGTAGTTGATGATCTTTTGTATA 1770
1819 GCAGAGAGGTTTATACAAAACTTCCAAATTTCACTGTTTATTTAGATGATCTTTATGACG 1878
1771 TCGGTGCTCAATPCGATGAATGACCAACCTGATTCATGTTGTTGAAAAATGGAATGTAG 1830
1879 CCATGGAATCTTAGAGATCTTAAGTTGTTACAGAACTAGTCAAAAGATGGATCTAT 1938
1831 ATGTGCAAGGATGTTGTTGAGACATGTTCCGATTTTATTTTATTTAGCAATTAAGATG 1890
1939 CACTAG--TGGACCAAAATGCCACAAACAAATGAAATATGTTTGTGGGTTTCTCAATA 1995
1891 CAATCTGTTGGATGAGATGAGCTTTTAAATGGCAAGCGCGATGTAACCTAGGCATG 1950
1996 CTTTAAATGATATAGCAAAAGAGAGCTGAGAGGAGAGGCGGATGTCTAGGCTACA 2055
1951 TTATTCAAACTTGGTTGGAACCTAATGAATAGTATGTTGAGAGAGCTATATGACCAAGAG 2010
2056 TTCAAATGTTTGGAAAGTCCAACTTGAAGCTTACACGAAAGAGAGAGATGCTGAAG 2115
2011 ATGCTTATGTCACCAATTAATGAATATATGGAACGCTTACGTCATTTTGCATTTAG 2070
2116 CTAATATGTCGCATCTTCAATGAATATACATAGAGATGCGATGTTGTCATAGCATTTG 2175
2071 GCCCGATTTG 2080
2176 GAACAGTCGT 2185

RESULT 10
US-09-315-861-1
; Sequence 1, Application US/09315861
; Patent No. 6114160
; GENERAL INFORMATION:
; APPLICANT: Rodney B. Croteau, Mark R. Wildung
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL

;/ TITLE OF INVENTION: BIOSYNTHESIS
;/ NUMBER OF SEQUENCES: 6
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Donald L. Stephens Jr.
;/ STREET: One World Trade Center
;/ STREET: 121 S.W. Salmon Street
;/ STREET: Suite 1600
;/ CITY: Portland
;/ STATE: Oregon
;/ COUNTRY: United States of America
;/ ZIP: 97204
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Disk, 3-1/2 inch
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: MS DOS
;/ SOFTWARE: WordPerfect 5.1
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/315,861
;/ FILING DATE:
;/ CLASSIFICATION:
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: 08/843,363
;/ FILING DATE:
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Donald L. Stephens Jr.
;/ REGISTRATION NUMBER: 34,022
;/ REFERENCE/DOCKET NUMBER: 4630-46842/DLS
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (503) 226-7391
;/ TELEFAX: (503) 228-9446
;/ INFORMATION FOR SEQ ID NO: 1:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 2700 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: double stranded
;/ TOPOLOGY: linear
;/ US-09-315-861-1

Query Match 5.5%; Score 153.8; DB 3; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;
Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;

QY 335 AAATGTAAGAAATTCGTTTCTTCAATGACACAGCATGGTAGCCATGTT-----CCC 388
Db 414 AGACGGAGATATCAGTCCGCTGTCATACGACATGCGTGGGTGGCGAGGCTGCGACCAT 473
QY 389 TTCTCCAAACTCACCAAAATCGCTTTTCCCTGAGTGTCTCAATTTGGTTAATTAATA 448
Db 474 TTCTCTGTATGATCTGGAAGCCACGGTTTCTCAGGCCCTCACTGGGTTTTCACAA 533
QY 449 TCAGCTTAATGATGGTTTCATGGGCTCTTGTAAATCACACTCATATAATCATATACCCGTT 508
Db 534 CCAGCTCCAGGATGATCGTGGGTATCGAATCGCACTTTAGTTTATGCGATCGATGCT 593
QY 509 GCTTAAAGATTCCTATCTTCAACATTAGCATGTATTTGTCATTAAAGATGGATGT 568
Db 594 -----TAACACGACCAATCTGTATCGCCCTCTCGGTTTGGAAAC 635
QY 569 TGGGGAAGATCAATAATAAAGTCTTAAGTTTATTTAGTCAATCTTCTTTCAGCTAC 628
Db 636 AGGCGACGCAAGTACACAGGTGCTGAGTTTATTCAGAGAACTAAGATTACTCAA 695
QY 629 TGAATAAAGTCAACCACTCTCCCATTTGTCATCATATTTCTGGTTTGTGATGATG 688
Db 696 TGAGGAAGATGAGTTGTTCTCC-----GGATTTCCAATAATCTTCTCTGCTCTGCTGCAAAA 752
QY 689 TGGAAAAACTTGGACATAAACCTCTTCAAAACAAACAGATTTTGTAGTTGATGCTACA 748
Db 753 GGCAAAAGCGTTGGGATCAATCTTCTTACGATCTTCCATTTATCAATATTGTCAC 812
QY 749 TAAGAGGGAATTGGAGCAAAAAGATGCCAT-----TCAAAATGAGATGGATGATCTT 802
Db 813 AACACGGAAGCCAGGCTTACAGATGTTTCTGGGCGACGACAAATATTCAGGCCAACAT 872

;; PRIOR FILING DATE: 1999-09-17
;; PRIOR APPLICATION NUMBER: 60/130,628
;; PRIOR FILING DATE: 1999-04-22
;; PRIOR APPLICATION NUMBER: 60/150,262
;; PRIOR FILING DATE: 1999-08-23
;; NUMBER OF SEQ ID NOS: 58
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 43
;; LENGTH: 2700
;; TYPE: DNA
;; ORGANISM: Taxus brevifolia
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (22)...(2607)
;; OTHER INFORMATION: taxadiene synthase
US-09-887-586A-43

Query Match 5.5%; Score 153.8; DB 4; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;
Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;

QY 335 AAATGTAGAAATTTCTGTTCTTCTCATATGACACACATGGGTAGCCATGGT-----CCC 388
DB 414 AGACGGAGATTCAGTCGGTCTGCAATACGACACTCGTGGGTGGGAGGCTGGCGACCAT 473
QY 389 TTCTCCAAACTCACCCAAATCGCCTTGTGTTTCCCTCAGTGTCTCAATTTGGTTAAATTAATAA 448
DB 474 TTCTCTGATGATCTGAGAGCCACCGTTTCTCAGGCCCTCACTGGGTTTTCACAA 533
QY 449 TCAGTTAATGATGTTTCATGCGGTCTGTTTAAATCACACTCATATCAATTAATCAACCCGTT 508
DB 534 CCAGCTCCAGGATGATCGTGGGTATCGAATCGCACCTTTAGTTTATGGATCGATTGCT 593
QY 509 GCTTAAAGATCTCTATCTTCAACATTAGCATGTTTGTGCAATTAAGAGATGAAATGT 568
DB 594 -----TAACAGCACCAATCTCTGTTATCGCCCTCTCGGTTTGGAAAC 635
QY 569 TGGGAAGATCAATAAATAAGGCTAAAGTTTATTTAGTCAATCTTGTCTCAGCTAC 628
DB 636 AGGCACAGCAAGTACAAAGGCTGCTGAGTTTATTCAGAGAACTAAGNTACTCA 695
QY 629 TGAAAAAGTCAACATCTCCCATTTGTTTGAATCAATTTCTGTTGTTGCTTGAGTA 688
DB 696 TGAGGAAGATGAGTTGTGCCCC--GGATTTCCAAATAATCTTCTGCTGCTGCTGCAAAA 752
QY 689 TGCAGAAAATTTGGACATAAACCTCTCTTCAAAACAAACAGATTTAGTTTGAATCTACA 748
DB 753 GGCAGAACGTTGGGATCAATCTTCTTACGATCTTCCATTTATCAAAATATTTGTCGAC 812
QY 749 TAAGAGGAATTTGGAGCAAAAAGATGCCAT-----TCAATAGAGATGGATGATCTT 802
DB 813 AACACGGGAAGCCAGGCTTACAGATGTTTCTCGGCGACGACAAATATTCAGGCCAACAT 872
QY 803 GCGGTATATCTCTGAGAGCTCGGTAATTTATATAGTTGGAATATGGTGAAGAAATATCA 862
DB 873 GTTGAATGCGTTGGAAGGCTCTCGAGGAAGTTATTGACTGGAAACAAGATTATGAGGTTTCA 932
QY 863 GATGAAAAATGTTCTGTTTTCACATCCATCAGCAACAGCTGCTGCTTTCATTATCA 922
DB 933 AAGTAAAGATGGAATCTTTCCTGAGCTCCCTCCCTCCACTGCTGTGATGATGAATAC 992
QY 923 TCAAAATCCTGGTTGTCTTAAATTTTAAATTTCACTTTTGACAAAGTTTGGTAATGCAAT 982
DB 993 AGGGAGCAAAAAATGTTTTCACCTTTCTCAACAACTGCTGACAAATTCGCGGCTGGCT 1052
QY 983 CCCAACAGTTTATCTCTGATGATTTATTTATCCGACTTCTATGTTGTCACAAATTTGAAG 1042
DB 1053 GCCCTGTATGATTTCCATCGATCTGCTGGAACGCTTTTCGCTGTTTGAATCAATTTGACA 1112
QY 1043 ATTAGGAATTTTACACCAATTTTTCAGAGTGAATTTTAAATGTTTGTAGTGAACATACAG 1102
DB 1113 TCTCGGAATCGGTGCGCAATTTTCAACAGAAATCAAGAGGAGCTCTTGTGATTTATGTTACAG 1172

QY 1103 ATGTTGGGTGGAAACGAG-----ATGAGCAAAATATTTCATGATGTTGTGAAC 1147
DB 1173 ACATTTGGAGTGAAGGGGCATCGGTTGGGGCAGAGACAGCCTTGTTCAGATCTCAACAC 1232
QY 1148 ATGTGCTTTAGCCTTTTCGGTTTATTAAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATT 1207
DB 1233 CACAGCCCTCGGCTCGGAACCTTCGCAATGACCGGATACAAATGTTTCTTCAGACGTTTT 1292
QY 1208 GGCTGAATTAATTAATGAATTAGCTTTTGAAGACGAATATGAGCTC-----1254
DB 1293 GAATAATTTCAAGATGAAACCGGGGTTCTTCTCTCGGGGCCAAACCCATGTGCA 1352
QY 1255 -TTGAAACATATCATCGGTCACATATATTATACCAAGA--GGATTTATCTTCTGGAAC 1311
DB 1353 ATTGAGAACGTTGGTGAATCTTTTTCAGAGCTTCCGACCTTGCATTTCTCTGAGCAAGAGC 1412
QY 1312 AAATCTTGAAGTCAGCTGATTTCTTCAAGAGATAATATCCACTGATTCACACAGGCTTT 1371
DB 1413 TATGACGATGCTAGAAATTTTCAGAACCATATCTTAGAGAGGCACTTGCACAGAAAT 1472
QY 1372 CTAAATTAATTC-----CAAGAGGTGAAATGCTCTTAAGTTCCTTATCAATAC 1423
DB 1473 CTCAACCAATACAAAACTATTTCAAAGAGATTGAGTACGTTGGTGGAGTACCTTGGCAAT 1532
QY 1424 CGGTTTAGAACGCATAAACAACCTAGACGAATATATACAGCTTTTACATGTAGCAATACAG 1483
DB 1533 GAGTATCCACGCTTAGAACGCAAGATTTATTGATTCATATGACGCAATATTATGTATG 1592
QY 1484 AATTCTGAAAACTACATATCACTCATCAAAATATTAGTAACTGATTAACCTTAAGGTGCGC 1543
DB 1593 GCAGAGGAAGACTCTATATAGATGCCATCTTTGATGTAATTCAAATGTTTAGAATTTGCG 1652
QY 1544 TGTTGAAGATTTTACACTGCCCAATCTATTATGTCGAAGATTTAAAGGCTTTGAAG 1603
DB 1653 AAAATTTGACTTCAATATCGTACAACTTTGCAATCAAGAGGATTTGAAGCTTCTAACAG 1712
QY 1604 GTGGTGTAGAGATAAGTTGGACCCAGCTCAAGTTTCTAGGCAAAAGACCGCTACTG 1663
DB 1713 ATGGTGAAGGAATCCGGCATGGCAGATATAAATTTCACTCGACACCGAGTGGCGAGGT 1772
QY 1664 TTAATTTCTCTGTTGTGCAACACTTTTCTCTCCCAATTTATCAGATGCGGCTATTTCATG 1723
DB 1773 TTAATTTTCACTCAGCTACA-----TTTGAACCCGAATTTCTGCCACTAGAAATTTGCC 1826
QY 1724 GGCACAAAATGGCATATTAACTACAGTATGTTGATGACTTTTGTATATCGGTTGATCAAT 1783
DB 1827 CACAAAAATTTGGTTGTTTACAACTCTTTTTCAGATATGCTGACATCTTTTGCACACT 1886
QY 1784 CGATGAATTTGACCAACCTGATTCATGTTGAAAAATGGAATGTAGATGTGCAACAGGA 1843
DB 1887 AGATGAATTTGAAAGTTTCACTGAGGGAGTAAAGAGATGGGATACATCTTTTGTACATGA 1946
QY 1844 TTGTTGTTTTCAGACATGTTTCGGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGAT 1903
DB 1947 --GATTTCCAGAGTATGCAAACTTCTGCTTTAAAGTTTGGTTTCAATTAATGGAAGAGT 2003
QY 1904 TGGAGATGAAGCTTTTAAATGGCAGCGCGATGTAACATAGCCATGTTATTTCAAACTTG 1963
DB 2004 AAATAATGATGTTGTTTAAAGGTACAAAGGACGTGACATGCTCGCTCACATAAGAAACCC 2063
QY 1964 GTTCGAACTAATGAATAGTATGTTGAGAGAGCTATATGGACAGAGATGCTTATGTGCC 2023
DB 2064 GGAGTTGATCTTCAATTTGTTATGTTACAGAAAGGGAGTGGCTTGAAGCCGGTATATACC 2123
QY 2024 AACATTAATTAATATATGAAAAACGCTTACGTTCTCATTTGCAATTAGGCCCGATTTGTCAA 2083
DB 2124 AACTTTTGAAGAGTACTTAAAGCTTATGCTATATCAGTAGGCCCTTGGACCGTGTACCT 2183
QY 2084 CCGGCTATTATCTTTGTTGGGGCCCAAAATATCAGAGAGATTTGTTGAAAGCTCTGAATA 2143
DB 2184 ACACCAATACTACTAATATGGTGGAGCTTGTGMAAGATGATGTTTCTTGTGAGAAAGTGCACTA 2243
QY 2144 TC---ATAATCTATTTAAGCTAATGACGACGCGGGTCTGAGCTTCTTAAACGATATCCATAG 2200

Db 2244 TCCCTCAATATGTTGAGCTGTATCTTGGCTGGGACTTAACAAACGACACCAAAAC 2303
QY 2201 CTTCAAGAGGGAATTAAGGAAGGCAAAATAAAGCGGTAGCAATGCAATTTAGTAACGG 2260
Db 2304 ATATCAGGCTGAAAGGCTCGAGGACAAACAGCCTCAGGCATAGCATGTATATGAAGGA 2363
QY 2261 AGAAAGTGGAAAGTGGAAAGAGAGGTTCTGGAGGAGAT 2299
Db 2364 TAAATCAGGAGCACTGAGGAAGATGCCATTAAGCAAT 2402

RESULT 13
US-09-895-752-43
; Sequence 43, Application US/09895752
; Patent No. 6559297
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65592971, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/895,752
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Taxus brevifolia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)....(2607)
; OTHER INFORMATION: taxadiene synthase
US-09-895-752-43

Query Match 5.5%; Score 153.8; DB 4; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;
Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;

QY 335 AAATGTAGAAATTTCTTTCTTCATATGACACAGCATGGGTAGCCATGGT-----CCC 388
Db 414 AGACGAGATATCAGTCCGTTCTGCAATACGACATCTGGTGGGTGGCGAGGCTGGCGACCAT 473
QY 389 TTCTCCAAACTCACCCAAATPGCCTTTGTTTCCCTGAGTGTCTCAATTTGGTTAATTAATAA 448
Db 474 TTCTCTGTATGGATCTGAGAAGCCACGGTTTCTCTCAGGCCCTCAACTGGGTTTTCAACAA 533
QY 449 TCAGCTTAATGATGGTTCATGGGTCTTGTGTAATACACACTCAATATCAATCAACCGGTT 508
Db 534 CCAGCTCCAGGATGGATCGTGGGGTATCGAATCGCACTTTAGTTTATGCGATCGATTGCT 593
QY 509 GCTTAAAGATTTCTATCTTCAACATTAGCATGTATTGTTGCAATTAATAAGATGGAATGT 568
Db 594 -----TAAACAGCAATTTCTGTTATCGCCCTCTCGGTTTGGAAAC 635
QY 569 TGGGGAAGATCAATAAATAAGGTCTAAGTTTATTTAGTCAAAATCTTCCTCAGCTAC 628
Db 636 AGGGCAAGCAAGTACAAAGGTGCTGAGTTTATTTGCGAGAAATCTAAGATTAATCTCAA 695
QY 629 TGAABAAGTCAACATCTCCATTTGTTGACATCATATTTCTCGTTTGTGTTGAGTA 688
Db 696 TGAGGAAGATGATGTGTGCTCC-----GGATTTCCAAATTAATTTCTTCTGCTCTGCTGCAAA 752

QY 689 TCGGAAAACTTTGGACATAAAACCTCTCTTTCAAAAACAAACAGATTTTAGTTGATGCTACA 748
Db 753 GSCAAAACGGTTGGGGATCAATCTTCTTACGATCTTCATTTATCAAAATATTTCGAC 812
QY 749 TAAAGAGGAATTTGGAGCAAAAAGATGCCAT-----TCAATGAGATGATGATGATCTT 802
Db 813 AACACGGGAAGCAGGCTTTACAGATGTTTCTCGGCGAGCAGACAATTTTCAGCCAAAT 872
QY 803 GCGGTATATCTCTGAAGGACTCGGTAAATTTATATATGATTTGGAATATGGTGAAGAAATATCA 862
Db 873 GTTGAATGCGTTGGAAGGCTTCGAGGAAGTTATGACTTGAACAAGATTTATGAGTTTCA 932
QY 863 GATGAAAAATGTTCTGTTTTCAACTCAACATCAGACAGCTGCTGCTTTTCAATTAATCA 922
Db 933 AAGTAAAGATGATCTTTCTGAGCTCCCTCCCTCACTGCTGTGTAATGATGATGATGAT 992
QY 923 TCAAAATCTGTTGTTCTTTAAATTTAAATTTCACTTTTGGACAGTTTGGTAAATGCACT 982
Db 993 AGGGGACGAAAAATGTTTCACCTTTCTCAACAATCTGCTCGACAATTCGGGCGCTGCC 1052
QY 983 CCCAACAGTTTATCTCTCATGATTTATTTATCCGACTTTCTATGCTTGACACAATTTGAAAG 1042
Db 1053 GCGCTGTATGATTCATCGATCTGCTGGAACGCTTTGCTGTTGATTAACATTTGAGCA 1112
QY 1043 ATTAGGAATTTCAACCAATTTGAGTGGAAATTTAAATTTTGTAGATGAACATACAG 1102
Db 1113 TCTCGAATCGGTCGCCATTTCAACACAGAAATCAAGGAGCTCTTGTATGTTGCTACAG 1172
QY 1103 ATGTTGGTGGAAACGAG-----ATGAGCAAAATTTATCATGATGTTGTAAC 1147
Db 1173 ACATTGGAGTGAAGGGGCAATCGGTTGGGCGAGACAGCCTTTGTCAGATCTCAACAC 1232
QY 1148 ATGTTGCTTTAGCCTTTTCGGTTATTTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATT 1207
Db 1233 CACAGCCTCGGCTCGGAATCTTTCGATGCAAGTACAAATGTTTCTTCAGACGTTTT 1292
QY 1208 GCGTGAATTTACTAATGAATTAGCTTTGAAAGCAAGATATGAGCTC-----1254
Db 1293 GAATAATTTCAAAAGATGAAAACGGGCGTTCTTCTCTCTCGGGGCAACCAATGTCGA 1352
QY 1255 -TTGAAACATATCATCGCTCACATATATTATATCAAGA--GGATTTATCTTCGGAAC 1311
Db 1353 ATTGAGAAGCGTGGTGAATCTTTTCAGAGCTTCGACCTTGCATTTCTTCGACAAAGAGC 1412
QY 1312 AAATCTTGAAGTCAGCTGATTTCTCAAAGAGATAATATCCACTGATTTCAACACAGGCTTT 1371
Db 1413 TATGAGCATGCTAGAAAATTTGCAGAACCAATCTTAGAGAGGCACTTTCGAACGAAAT 1472
QY 1372 CTAAATTAATCA-----CAAAGGTGGAATGCTTTAAGTTCCCTATCATATC 1423
Db 1473 CTCAACCAATACAAAACCTATTCAAAGAGATTTAGTACGTTGGTGGAGTACCCTTTGGC 1532
QY 1424 CGGTTTAGAACCATTAACACATAGACGAATATACAGCTTTTACAATGTAGACAATAACAAG 1483
Db 1533 GAGTATCCCAAGCTTTAGAACCGAAGTTATATGATTTATGATGACGAATTTATGATG 1592
QY 1484 AATTTCTGAAAACCTACATATCACTCATCAATATTTAGTAACACTGATTTACCTAAGTTGGC 1543
Db 1593 GCAGAGGAAGACTCTATATAGATGCACTCTTTTGAATTAATTTCAAAATGTTTAGAATGGC 1652
QY 1544 TGTGTAAGATTTCTACACCTGCCAATCTATTTATCGTGAAGATTTAAAGGTTCTTGAAG 1603
Db 1653 AAAATTTGGAATCTCAATATCTGTAATTTTGGCAATCAAGAGGATTTGAAGCTTTTAA 1712
QY 1604 GTGGGTGTAGAAATAAGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTG 1663
Db 1713 ATGTTGAAGGAATCCGGCATGCGAGATTAATTTCACTTCGACACCGAGTGGCGAGGT 1772
QY 1664 TTAATTTCTGTTGTCGAACACTTTGCTCTCCGAATTTATGAGTGGCGGTTTTCATG 1723
Db 1773 TTAATTTCTCATCAGCTACA-----TTTGAACCGGAATATTCTGCCCATGAAATTTGCT 1826
QY 1724 GGCACAAAATGGCATTTAACTACGATGTTGATGACTTTTTTTGATATCGTGGTACAAT 1783

Db 1827 CACAAAAATGGTGTGTTTCAAGTCTCTTTTATGATGATATGGCTGACATCTTTGCAACT 1886
QY 1784 CGATGAATGACCAACTGATCAATGCTGTGAAAAATGGAATGTAGATGTGCAAGGA 1843
Db 1887 AGATGAATGAAAGTTTCACTAGGAGGATGAAGAGATGGATACATCTTTGCTACATGA 1946
QY 1844 TTGTTGTTTCAGAGCATGTTCCGATTTTATTTTATAGCATTAAGAGATGCAATCTGTTGGAT 1903
Db 1947 --GATTCAGAGTGTATGCAAACTTGTCTTTAAAGTTTGGTTCAATTAATGGAAGAT 2003
QY 1904 TGGAGATGAAGCTTTTAAATGGCAGCGCGCATGTAACCTAGTTTATTCAAACTTG 1963
Db 2004 AAATAATGATGTGGTTAAGGTACAAGGACGTGACATGCTCGCTCACATAAGAAAAACCTG 2063
QY 1964 GTTGAACATAATGAATAGTATGTGAGAGAGCTATATGCAAGAGATGCTTATGTGCC 2023
Db 2064 GGAGTTGTACTTCAATGTTATGTACAGAAAGGAGTGGCTTGAAGCGGGTATATACC 2123
QY 2024 AACATTAATGAATATATGAAACGCTTACGTGTCAATTTGCAATAGGCCCGATTTGCA 2083
Db 2124 AACTTTGAAGAGTACTTAAAGACTTATGCTATATCAGTAGGCCTTGGACCGTGTACCT 2183
QY 2084 GCCGCTATTACTTTGTTGGGCCCAATATATCAGAGGATGTTTGAAGCTCTGAATA 2143
Db 2184 ACAACCAATACTACTAAATGGGTGAGCTTGTGAAAGATGATGTTGTGAGAAAGTGCACTA 2243
QY 2144 TC--ATAATCTATTAAAGTAAATGAGCAGCAGCGTGCAGCTTCTAAACGATATCCATAG 2200
Db 2244 TCCCTCAATATGTTGAGCTGTGATCTCTTGAGTGGCGACTTAAACAGCAGACCAAAAC 2303
QY 2201 CTTCAAGAGGAATTAAGGAAGGCAATTAACCGGTAGCATTTGCAATTTGAGTAACGG 2260
Db 2304 ATATCAGGCTGAAAGGCTCGAGGACAACAGCCTCAGGCATAGCATGCTATATGAAGA 2363
QY 2261 AGAAGTGGGAAGTGAAGAGAGGTTGTGGAGAGAT 2299
Db 2364 TAATCCAGGAGCAACTGAGGAAGATGCCATTAAGCACAT 2402

RESULT 14

US-09-903-012B-43
; Sequence 43, Application US/09903012B
; Patent No. 6569656
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65696561, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Mauna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/903,012B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ. ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Taxis brevifolia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)...(2607)
; OTHER INFORMATION: taxadiene synthase
US-09-903-012B-43

Query Match 5.5%; Score 153.8; DB 4; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;

Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;
QY 335 AAATGTAGAAATTTCTGTTTCTTATATGACACAGATGGGTAGCCATGGT-----CCC 388
Db 414 AGACGGAGATATCACTCGTCTGTCATACGACACTGGTGGGTGGCGAGCTGGCGACCAT 473
QY 389 TTCTCCAAACTCACCCTTGGTTTCCCTGAGTGTCTCAATTTGGTTAAATTAATAA 448
Db 474 TTCTCTGATGATCTGAGAGCCACGGTTTCTCAGGCCCTCACTGGGTTTCAACAA 533
QY 449 TCAGCTTAATGATGTTTCATGGGTTTAAATCACAATCATATATCAATCAATCAACCGTT 508
Db 534 CCAGCTCCAGGATGATCGTGGGTTATCGAATGACACTTTAGTTTATGCGATCGATTGCT 593
QY 509 GCTTAAAGATTTCTATCTTCAACANTAGCATGTTTTCGANTPAAAAAGATGGAATGT 568
Db 594 -----TAAACAGCAAAATCTGTTATCGCCCTCTCGGTTTGGAAAC 635
QY 569 TGGGGAAGATCAAAATAAAGTCTAAGTTTATTTAGTCAAAATCTTCTGCTGAGTA 628
Db 636 AGGCGACAGCCAAATGACAAAGGTTGCTGAGTTTATGAGAGAAATCTAAGATTACTCAA 695
QY 629 TCAAAAAAGTCAACCATCTCCCATTTGGTTTGAACATCATATTTCTGTTGCTGAGTA 688
Db 696 TGAGGAAGATGATTTGTCCTCC-----GGATTTCCAAATAATCTTCTGCTGCTGCAAAA 752
QY 689 TCGAAAACTTGGACATAAACCTCTTTTCAAAACAAACAGATTTTGTGTTGATGCTACA 748
Db 753 GCGAAAAAGCTTGGGGATCAATCTTCTTAGCATCTTCCATTTATCAATAATTTGTCGAC 812
QY 749 TAAGAGGGAATTTGGAGCAAAAAAGATGCCAT-----TCAATGAGATGGTATGATCT 802
Db 813 AACCGGGAAGCCAGGCTTACAGATTTTCTGGGCGACGACAAATATTCAGGCCAACAT 872
QY 803 GCGGTATATCTCTGAGGACTCGGTAAATTTATGATGGAATATGGTGAAGAAATATCA 862
Db 873 GTTGAATGCGTTGGAAGGTCTCGAGGAATTTATGACTGGAAACAGATTTAGGTTTCA 932
QY 863 GATGAAAAATGTTCTGTTTTCATCTCACCATCAGCAACAGCTGCTGCTTTTCAATTAATCA 922
Db 933 AAGTAAAGATGGAATCTTCTGAGTCCCTGCTCCACTGCTGCTGTTGATGATGAATAC 992
QY 923 TCAAAATCCCTGTTGCTTAAATTTAAATTTCACTTTTGGACAGTTTGGTAATGAGT 982
Db 993 AGGGAACGAAATGTTTCACTTTTCAACAAATCTGCTGACAAATTCGGCGGCTGCGT 1052
QY 983 CCCAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGTTTGCACAAATTTGAAG 1042
Db 1053 GCCCTGTATGTTTCCATCGATCTGCTGGAAGGCTTTGCTGTTGATACATTTAGCA 1112
QY 1043 ATTAGGAATTTCAACCATTTGAGAGTGGAAATTAATAATTTTATAGATAAATACAG 1102
Db 1113 TCTCGAATCGGTGCGCAATTTCAAAACAGAAATCAAAAGAGCTCTTGATTATGCTACAG 1172
QY 1103 ATGTTGGGTGGAACGAG-----ATGACAAATATTTATGATGTTTGAAC 1147
Db 1173 ACATTGAGTGAAGAGGGGCTCGGTTGGGCGAGACAGCGCTTGTCCAGATCTCAACAC 1232
QY 1148 ATGTGCTTTAGCCCTTTGCTTATTAAGGATCAATGGGTATGAAGTTTCCCAGATCCATT 1207
Db 1233 CACAGCCCTCGGCTCGAACTCTTCGATGCGAGATACAATGTTTCTTCAGAGCTTTT 1292
QY 1208 GGTGAAATTAATGAATTAGCTTTGAAAGACGAATATGCAAGCT-----1254
Db 1293 GAATAATTTCAAAGATGAAACCGGGGTTCTTCTCTCTGCGGCGCAACCCATGTCGA 1352
QY 1255 -TTGAAACATATCATCGTGCACATATATATACAGA--GGATTATCTTCTGGAAGAAC 1311
Db 1353 ATTGAGAAAGCGTGGTGAATCTTTTCAGAGCTTCGACCTTCGATTTCTGACGAAAGAGC 1412
QY 1312 AAATCTTTGAGTCAAGTGTATTTCTCTCAAGAGATATATCCACTGATTTCAACAGGCTTT 1371
Db 1413 TATGGAGATGCTAGAAAATTTGAGAACCAATATCTTAGAGGACACTTCCACAGAAAT 1472

BIOSYNTHESIS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Donald L. Stephens Jr.
STREET: One World Trade Center
121 S.W. Salmon Street
Suite 1600
CITY: Portland
STATE: Oregon
COUNTRY: United States of America
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Disk, 3-1/2 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA: US/09/593,253
FILING DATE: 13-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER:
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Donald L. Stephens Jr.
REGISTRATION NUMBER: 34,022
REFERENCE/DOCKET NUMBER: 4630-46842/DLS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 226-7391
TELEFAX: (503) 228-9446
INFORMATION FOR SEQ ID NO: 1
SEQUENCE CHARACTERISTICS:
LENGTH: 2700 base pairs
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-593-253-1
Query Match 5.5%; Score 153.8; DB 4; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;
Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;
QY 335 AAATGTAGAAAATTCCTGTTTCTTATGACACAGCATGGTAGCATGGT-----CCC 388
Db 414 AGACGGAGATATCAGTCCGCTCTGATACGACATCGGTGGGTGGCGAGGCTGGCGACCAT 473
QY 389 TTCTCCAAACTACCCAAATCGCCTGTTTCCCTGAGTGTCTCAATTGGTTAATTATAA 448
Db 474 TTCTCTGATGGATCTGAGAAGCCACGGTTTCTCAGGCGCTCACTGGGTTTTCACAA 533
QY 449 TCAGCTTAATGATGGTTCATGGGTCTTGTAAATCACACTCATAATCAATCAATCAATCAAT 508
Db 534 CGAGCTCCAGATGGATCGTGGGTATCGAATCGCACTTTAGTTATGCGATCGATTGCT 593
QY 509 GCTTAAAGATTCCTATCTTCAACATTAGCATGTTGGTATTAAGAGATGGAATGT 568
Db 594 -----TAAACACGCAATTCGTATCGCCCTCTCGGTTTGGAAAC 635
QY 569 TGGGGAAGATCAATAATAAGCTTAAGTTTATGATCAAACTCTGCTTCAGCTAC 628
Db 636 AGGCGCAGCAAGTACACAGGTGCTGAGTTTATTCGAGAGATCTAAGATTACTCAA 695
QY 629 TGAAGAAAGTCAACCATCTCCATTTGGTTTGGATCATATTTCTGCTGCTGCTGCTGCTGCT 688
Db 696 TGAGGAGATGAGTTGCTCCCT---GGATTTCCAAATATCTTCTGCTGCTGCTGCTGCTGCT 752
QY 689 TGCGAAAGAACTTGGACATAAACCCTCTTCAAAACAAACAGATTGTTGTTGTTGTTGTTGTT 748
Db 753 GGCAGGAGGCTTGGGATCAATCTCTTACGATCTTCCATTTTATCAAAATATTTGTCGAC 812
QY 749 TAAGAGGGAATGGAGCAAAAGATGCCAT-----TCAATGAGATGGATGATGATCTT 802

QY 1372 CTAATTAATTCA-----CAAGAGGTGAAATGCTCTTAAGTTCCCTTATCAATAC 1423
Db 1473 CTCAACCAATACAAACTATTCAAGAGATTGAGTACGTGGGAGTACCTTTGGCAGAT 1532
QY 1424 CGGTTTGAAGCAGATAAAGTACATAGCAAGTATACAGCTTTTCAATGTAGACAAATACAG 1483
Db 1533 GAGTATCCACCTTAGAGCCAGAGTTATGATTCAATGACGCAATTAATGATG 1592
QY 1484 AATTCGAAACTACATCACTCAATTAATTAATGATCACTGATTAATGATGAGTTGGC 1543
Db 1593 CGAGAGGAGACTCTATATAGAAATGCACTTTGAGTAATCAAAATGTTTGAATGGC 1652
QY 1544 TGTTGAAGTTTCTACACTGCACTCAATCTTATCTGAGAGATTAAGAGTCTTGAAG 1603
Db 1653 AAAATTGGACITCAATATCGTCAATCTTTGATCAAGAGGAGTTGAAGCTTCAACAG 1712
QY 1604 GTGGGTGGTAGAATAAGTGGACCGAGTCAAGTTTGTAGGCAAAAGACCGCTACTG 1663
Db 1713 ATGGTGAAGGATCGGCAATGGCAGATATAATTTCACTCGACACCGAGTGGGAGGT 1772
QY 1664 TTATTTCTGTTGTCGCAACTTTTCTGCTCCGAAATTAATGATGCGGATTTTCATG 1723
Db 1773 TTATTTTTCATCAGCTACA-----TTTGAACCGCAATATTTCTGCCACTAGAAATGCTT 1826
QY 1724 GGCAGAAATGCAATTAATTAACATGATGATGATGATGATGATGATGATGATGATGATGAT 1783
Db 1827 CACAAATATGTTGTTTCAAGTCTTTTGAATGATGATGATGATGATGATGATGATGATGAT 1886
QY 1784 CGATGAATTCACCAACTGATTCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1843
Db 1887 AGATGAATTAAGAGTTTCACTGAGGAGTAAAGATGGAATGATGATGATGATGATGATGATGAT 1946
QY 1844 TTGTTTTCAGAGCATGTTGCGAATTTATTTTAGCATTAAGAGTGAATCTGTTGGAT 1903
Db 1947 --GATTCAGAGTATGCAACTGTTTAAAGTTTGGTTCAAAATTAATGGAAGAAT 2003
QY 1904 TGGAGATGAGCTTTTAAATGGCAAGCGCGAGTAACTAGCATGTTTATCAAACTTG 1963
Db 2004 AATAATGATGTTGGTTAAGTACAGGACGTGATGATGATGATGATGATGATGATGATGATGAT 2063
QY 1964 GTTGAAGTAAATGAATGATGTTGAGAGAGCTATATGACAGAGATGCTTTATGTCC 2023
Db 2064 GGAATGTTACTTCAATGTTATGACAGAAAGGAGTGGCTTGAAGCGGGTATATACC 2123
QY 2024 AACTTAATGAATATGAAAGCGTTACGTGTCATTTGATGATGATGATGATGATGATGATGATGAT 2083
Db 2124 AACTTTTGAAGTACTTAAAGACTTATGCTATATCAGTAGGCGCTTGGACCGGTGTACCCCT 2183
QY 2084 GCGGCTATTTACTTTGTGGGGCCCAATTTATCAGAGGAGATGTTGMAAGCTCTGAATA 2143
Db 2184 ACAACAACTACTAATGAGTGGTGGCTTGTGAAGATGATGTTGTTGAGAAAGTGCATTA 2243
QY 2144 TC---ATAATCTTAAAGCTAATGACACGCGGGTGCATCTTCTAAACGATATCCATAG 2200
Db 2244 TCCCTCAATATCTTTGAGCTTGTATCTTTGAGCTGCGGACTAACCAACGACACCAAAAC 2303
QY 2201 CTTCAAGAGGAATTTAAGAGGCAATTAAGCGGTGAGCATTTGATTTGATGATGATGATGATGATGAT 2260
Db 2304 ATATCAGGCTGAAAAGGCTCGAGGACCAACAAAGCTCAGGCAATGATGATGATGATGATGATGAT 2363
QY 2261 AGAAGTGGGAAATGGAAGAGGTTGTGGAGGAGAT 2299
Db 2364 TAATCCAGGAGCACTGAGGAGATGCCATTAAGCAT 2402

RESULT 15
US-09-593-253-1
; Sequence 1, Application US/09593253
; Patent No. 6610527
; GENERAL INFORMATION:
; APPLICANT: Rodney B. Croteau, Mark R. Wildung
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL

Db 813 AACACGGGAAGCCAGCTTTACAGATGTTTTCTGCGGACGACAGACAATATTCAGCCAAACAT 872
QY 803 GCGGTATATCTCTGAAGACTCGGTAAATTTATATGATTCGAATATGCTGAAGAAATATCA 862
Db 873 GTTGAATGCGTTGGAAGTCTCGAGGAAGTATGACTGGAACAGAGATATGAGGTTTCA 932
QY 863 GATGAAAAATGTTCTGTTTTCAACTCACCATCAGCAACAGCTGCTGCTTTCAATTAATCA 922
Db 933 AAGTAAAGATGATCTTTCTGAGCTCCCTGCCCTCCACTGCTGTGTAATGATATGATAC 992
QY 923 TCAGAAATCTGCTGCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 982
Db 993 AGGGACGAAAAATGTTTCACTTTTCTCAACAATCTGCTCGACAAAATTCGGCGGCTGCT 1052
QY 983 CCCACAGTTTTATCTCATGATTTATTTATCGACTTTCTATGTTTGCACAAATTCGAAG 1042
Db 1053 GCCCTGATGATTCATCGACTGCTGGAACGCTTTTCGCTGTTGATAACATTTAGACA 1112
QY 1043 ATTAGAAATTCACACCTTTTCAGAGTGGAAATTAATAATGTTTTAGATGAACAATACAG 1102
Db 1113 TCTCGAAATCGGTCGCCATTTCAACAAGAAATCAAGGAGCTCTTGATTAATGCTACAG 1172
QY 1103 ATGTTGGGTGGAACAG-----ATGAGCAAAATATTCATGATGATGTTGTAAC 1147
Db 1173 ACATTTGAGTGAAGGGGACGCGTTGGGCGACAGACGCTTTGTTCCAGATCTCAACAC 1232
QY 1148 ATGCTCTTTAGCCTTTTCGTTTATTAAGGATCAATCGGTATGAAGTTTCCCGACATCCATT 1207
Db 1233 CACAGCCCTCGGCTCGCACTCTTCGCATGCAACGATACATGTTTCTTCAGAGCTTTT 1292
QY 1208 GGCTGAATTAATTAATGATGCTTTTGAAGAAGAAATATGACGCTC----- 1254
Db 1293 GAATAAATTTCAAGATGAAGAAACGGGCGGTTCTTCTCTCTCGGGCCAAACCCATGTCGA 1352
QY 1255 -TTGAACATATCATGCTGCATATATATATACCAAGA--GGATTTATCTCTGGAAGAAC 1311
Db 1353 ATTGAGAAGCGTGGTGAATCTTTTCAGAGCTTCGACCTTGCACTTTCTGACGAAGAGC 1412
QY 1312 AAATCTTGAAGTCACTGATTTCTTCAAGAGATAATATCACTGATTCATCAACAGGCTTT 1371
Db 1413 TATGACGATGCTAGAAAATTTGAGAACCAATCTTACAGAGGCACTTGCAACGAAAT 1472
QY 1372 CTAAATTAATTC-----CAAGAGGTGGAAAAATGCTTTAAGTTCCCTATCAATAC 1423
Db 1473 CTCACCAATACAAAATATTTCAAGAGATGAGTACGTGTGGATACCTTGGCCAT 1532
QY 1424 CGGTTTAGACGATACAACTAGACGAAATATACAGCTTTACAAATGTAGACATACAG 1483
Db 1533 GAGTATCCCGCTTAGAAGCCAGAGTTATATTGATTCATATGACGACAAATATGATG 1592
QY 1484 AATTTGAAAACATACATCACTCATCAAAATATTAGTAACTGATTAACCTAAGGTTGGC 1543
Db 1593 GCAGAGGAAGACTCTATATAGAAATGCCATCTTTGAGTAAATTCAAATGTTTGAATTTGC 1652
QY 1544 TGTTGAAGATTTTACACCTGCCAATCTATTTATGCTGAAGAAATTAAGGCTTTGAAAG 1603
Db 1653 AAAATTTGACTTCAATATCGTACAACTTTTGCATCAAGAGAGTTGAAGCTTTCTAACAG 1712
QY 1604 GTGGGTGTAGAGAATAAGTTGACCGACCTCAAGTTTGTAGGCAAGACCGCTACTG 1663
Db 1713 ATGTTGAAGGAATCCGGCATGAGAGATATAAATTTTCACTCGACCCGAGTGGCGGAGGT 1772
QY 1664 TTAATTTCTGTTGCTGCAACACATTTTGTCTCCGAAATATATCATGATGCGGTATTTCTATG 1723
Db 1773 TTAATTTTTCATCACTACA-----TTTGAACCCGAAATATTTCTGCCACTGAAATGCTT 1826
QY 1724 GGCACAAAATGGCAATATACTACAGATGTTGATGACTTTTTTGTATATCGTGGTACAAT 1783
Db 1827 CACAAAAAATTTGTTGTTTAAAGTCTTTTGTATGATATGCTGACATCTTTGCAACACT 1886
QY 1784 CGATGAATTCACCACTGATTAATGTTGAAAAATGGAATGGAATGATGATGTCGACAGGA 1843
Db 1887 AGATGAATTAAGATTTCACTGAGGGAGTAAAGAGATGGGATACATCTTTTGTCTACATGA 1946

QY 1844 TTGTTGTTCCAGAGCATGTTTCGGATTTTATTTTATAGCATTTAAAGATGCAATCTGTTGGAT 1903
Db 1947 ---GATTCAGAGTGTATGCAAACTTGTCTTAAAGTTTGGTTCAAATTAATGGAAGAAT 2003
QY 1904 TGGAGATGAAGCTTTTAAATGCGAAGCGCGCATCTAACTAGCCATGTTTATTCAAACTTG 1963
Db 2004 AATTAATGATGTTGTTTAAAGTTTAAAGGACGTGACATGCTTCGCTCACAATAAGAAAAACCTG 2063
QY 1964 GTTGAACCTAATTAATGATGATGTTTGAAGAACTATATGGAACAAGATGCTTATGTGCC 2023
Db 2064 GGAGTTGTACTTCAATTTGTTATGTAACAAGAGGAGTGGCTTTGAAGCCGGTATATACC 2123
QY 2024 AACATTAATGATGATGTAATGGAACCGCTTACGTTGCTATTTGCAATTTAGGCCCGCATTTGTCAA 2083
Db 2124 AACTTTTGAAGAGTACTTTAAAGACTTTATGCTATATCAGTAGGCTTTGGACCGTGTACCT 2183
QY 2084 GCGGCTTATTTACTTTTGTGGGSCCAAAATTTATCAGAGAGATTTGTTGAAGCTCTGAATA 2143
Db 2184 ACAACCAATFACTACTTAATGGGTGAGCTTGTGAAGATGATGTTGTGTGAGAAAGTGCCTA 2243
QY 2144 TC---ATAATCTATTTAAGCTAATGAGCACGCGAGGTGCACTTTCTAAACGATATCCATAG 2200
Db 2244 TCCCTCAAATATGTTTGAAGCTTGTATCCTTGAGCTGGCGACTAACAAAGCACCAAAAC 2303
QY 2201 CTTCAAGAGGGAATTTAAGGAAGGCAATTAACCGGGTAGCATTTGCAATTTGAGTAAACGG 2260
Db 2304 ATATCAGGCTGAAAAGGCTCGAGGACACAAAGCCTCAGGCATAGCATGCTATATGAAGGA 2363
QY 2261 AGAAAAGTGGGAAAGTGGAAAGAGAGGTTTGTGGAGAGAT 2299
Db 2364 TAAATCCAGGACCACTGAGGAAGATGCCATTAAGCAT 2402

Search completed: October 17, 2004, 03:30:43
Job time : 238.559 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 16:35:39 ; Search time 1304.77 Seconds
(without alignments)
11232.937 Million cell updates/sec

Title: US-10-041-018-361
Perfect score: 2792
Sequence: 1 cccactcctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 2792 | 100.0 | 2792 | 12 | Adm98941 Diterpene |
| 2 | 2582.6 | 92.5 | 3117 | 12 | Adm98942 Diterpene |
| 3 | 1428.6 | 51.2 | 2594 | 12 | Adm98948 Diterpene |
| 4 | 809.4 | 29.0 | 2638 | 12 | Adm98947 Diterpene |
| 5 | 782.2 | 28.0 | 2658 | 12 | Adm98958 Diterpene |
| 6 | 782.2 | 28.0 | 2658 | 12 | Adm98958 Diterpene |
| 7 | 780.6 | 28.0 | 2658 | 2 | Adm98961 Diterpene |
| 8 | 732.4 | 26.2 | 2506 | 12 | Aat60339 Pumpkin e |
| 9 | 453.2 | 16.2 | 2403 | 12 | Adm98957 Diterpene |
| 10 | 416.8 | 14.9 | 2154 | 8 | Adm94223 Corn Ent- |
| 11 | 379.4 | 13.6 | 2223 | 8 | Ada70130 Rice Gene |
| 12 | 379.4 | 13.6 | 2223 | 12 | Ada69971 Rice Gene |
| 13 | 330 | 11.8 | 1559 | 12 | Adi45176 Rice isop |
| 14 | 329.4 | 11.8 | 1875 | 8 | Adm98955 Diterpene |
| 15 | 315.6 | 11.3 | 2193 | 8 | Ada70068 Rice Gene |
| 16 | 241.6 | 8.7 | 2313 | 8 | Ada69972 Rice Gene |
| 17 | 238.2 | 8.5 | 1157 | 12 | Ada70298 Rice Gene |
| 18 | 194.8 | 7.0 | 2861 | 3 | Adm94233 Corn Ent- |
| 19 | 194.8 | 7.0 | 2861 | 12 | Aaa38937 Grand fir |
| 20 | 172.8 | 6.2 | 439 | 12 | Adm98945 Diterpene |
| 21 | 172 | 6.2 | 577 | 12 | Adm98953 Diterpene |
| | | | | | Adi45560 Wheat iso |

| | | | | | |
|----|-------|-----|-------|----|---------------------|
| 22 | 153.8 | 5.5 | 2700 | 2 | Aat97447 Pacific Y |
| 23 | 153.8 | 5.5 | 2700 | 3 | Aaa38931 Yew taxad |
| 24 | 153.8 | 5.5 | 2700 | 12 | Adm98943 Diterpene |
| 25 | 139.2 | 5.0 | 531 | 12 | Adm94227 Wheat Ent |
| 26 | 133.8 | 4.8 | 2705 | 12 | Adm98977 Diterpene |
| 27 | 130.6 | 4.7 | 468 | 12 | Adm94225 Rice Ent- |
| 28 | 130 | 4.7 | 339 | 12 | Adm97821 Maize ent |
| 29 | 128.4 | 4.6 | 314 | 12 | Adm97820 Maize ent |
| 30 | 120.8 | 4.3 | 2528 | 2 | Aax87531 B-alpha-b |
| 31 | 119.2 | 4.3 | 2424 | 2 | Aax87530 E-alpha-b |
| 32 | 119.2 | 4.3 | 2424 | 3 | Aaa38932 Grand fir |
| 33 | 119.2 | 4.3 | 2424 | 4 | Aaf73382 Grand fir |
| 34 | 119.2 | 4.3 | 2525 | 2 | Aax87530 E-alpha-b |
| 35 | 119.2 | 4.3 | 2525 | 2 | Aax87529 Grand fir |
| 36 | 119.2 | 4.3 | 2571 | 2 | Aax87504 Grand fir |
| 37 | 119.2 | 4.3 | 80374 | 12 | Adm98956 Diterpene |
| 38 | 119 | 4.3 | 632 | 5 | Aah87659 Peppermin |
| 39 | 118.4 | 4.2 | 426 | 12 | Adm97823 Maize ent |
| 40 | 116.8 | 4.2 | 463 | 12 | Adm98949 Diterpene |
| 41 | 113.4 | 4.1 | 2460 | 3 | Aaz299415 Spybean c |
| 42 | 113.4 | 4.1 | 2460 | 6 | Aad40203 Soybean c |
| 43 | 112.4 | 4.0 | 441 | 12 | Adm97824 Maize ent |
| 44 | 100.6 | 3.6 | 2543 | 12 | Adm94221 Soybean E |
| 45 | 99.8 | 3.6 | 230 | 12 | Adm97814 Maize ent |

ALIGNMENTS

RESULT 1
ADM98941
ID ADM98941 standard; DNA; 2792 BP.
XX
AC ADM98941;
XX
DT 01-JUL-2004 (first entry)
XX
DE Diterpene synthase DNA #1.
XX
KW Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
KW diterpene precursor; diterpene synthase; defence toxin;
KW volatile defensive signal; pollinator attractant; photoprotectant.
XX
OS Stevia rebaudiana.
XX
FN US2004072323-A1.
XX
PD 15-APR-2004.
XX
PF 07-JAN-2002; 2002US-00041018.
XX
PR 05-JAN-2001; 2001US-0259880P.
XX
PA (MATS/) MATSUDA S P T.
PA (HART/) HART E A.
XX
PI Matsuda SPT, Hart EA;
XX
DR WPI; 2004-373921/35.
XX
PT New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
PS Claim 3; SEQ ID NO 361; 38pp; English.
XX
CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene

synthase comprising growing several cells in the presence of a polyaromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents DNA encoding a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 2792 BP; 911 A; 474 C; 550 G; 857 T; 0 U; 0 Other;

Query Match 100.0%; Score 2792; DB 12; Length 2792;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2792; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCACTCATCCTTATCAACCAATACCATCGTTCTGCCACCGAGAGACTGATACGCGACG 60
DB 1 CCCACTCATCCTTATCAACCAATACCATCGTTCTGCCACCGAGAGACTGATACGCGACG 60

QY 61 GACCTGACGACGCTTTAATCTCTGTTGGACCACTTAAACAAATTCGACCATCATCT 120
DB 61 GACCTGACGACGCTTTAATCTCTGTTGGACCACTTAAACAAATTCGACCATCATCT 120

QY 121 ACTCCGGTGACGCTTTCAATTTAGTAAATCAGCAATCAATCAATCAATCAATCAAT 180
DB 121 ACTCCGGTGACGCTTTCAATTTAGTAAATCAGCAATCAATCAATCAATCAATCAAT 180

QY 181 TTTCACTATGATCGCGTCCCTTTGTTAAACCAATCAATCAATCAATCAATCAATCAAT 240
DB 181 TTTCACTATGATCGCGTCCCTTTGTTAAACCAATCAATCAATCAATCAATCAATCAAT 240

QY 241 CTATTCATACAGCATCAACTTCAATGTTGACCAATCAATCAATCAATCAATCAATCAAT 300
DB 241 CTATTCATACAGCATCAACTTCAATGTTGACCAATCAATCAATCAATCAATCAATCAAT 300

QY 301 ATACACCAAGACGCTTCAATGTTGACCAATCAATCAATCAATCAATCAATCAATCAAT 360
DB 301 ATACACCAAGACGCTTCAATGTTGACCAATCAATCAATCAATCAATCAATCAATCAAT 360

QY 361 ATGACACAGCATCGGTAGCATGCTTCTCCAAATCAATCAATCAATCAATCAATCAAT 420
DB 361 ATGACACAGCATCGGTAGCATGCTTCTCCAAATCAATCAATCAATCAATCAATCAAT 420

QY 421 CTGAGTGCTCAATTTGTTAATTAATCAATCAATCAATCAATCAATCAATCAATCAAT 480
DB 421 CTGAGTGCTCAATTTGTTAATTAATCAATCAATCAATCAATCAATCAATCAATCAAT 480

QY 481 ATCACTCATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 540
DB 481 ATCACTCATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 540

QY 541 GTATTGTTGCAATTAATAAGATGGAATGTTGGGGAAGATCAATTAATAAAGGTTCAAGTT 600
DB 541 GTATTGTTGCAATTAATAAGATGGAATGTTGGGGAAGATCAATTAATAAAGGTTCAAGTT 600

QY 601 TTATTGATCAATCTTCTGCTACCTTCAATTAATCAATCAATCAATCAATCAATCAAT 660
DB 601 TTATTGATCAATCTTCTGCTACCTTCAATTAATCAATCAATCAATCAATCAATCAAT 660

QY 661 ACATCATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 ACATCATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720

QY 721 AACCAACAGATTTTAGTTTGTGCTACATAGAGGGAATTTGGAGCAAAAAGATGCCATT 780

721 AACCAACAGATTTTAGTTTGTGCTACATAGAGGGAATTTGGAGCAAAAAGATGCCATT 780
781 CAATAGATGAGTGGATGATCTTCTGAGAGGACTCGTAAATTTATATGATT 840
791 CAATAGATGAGTGGATGATCTTCTGAGAGGACTCGTAAATTTATATGATT 840
841 GGAATATGTTGAGAGAAATATCAGATGAAAAATGGTCTGTTTCACTCCACCTCAGCA 900
841 GGAATATGTTGAGAGAAATATCAGATGAAAAATGGTCTGTTTCACTCCACCTCAGCA 900
901 CAGCTGCTGCTTCAATTAATCATCAAAATCCTGGTGTCTTAATTTAAATTCACATTT 960
901 CAGCTGCTGCTTCAATTAATCATCAAAATCCTGGTGTCTTAATTTAAATTCACATTT 960
961 TGGCAAGATTTGGTAAATGAGTCCCAACAGATTTATCCTCATGATTTATTTATCGACTTT 1020
961 TGGCAAGATTTGGTAAATGAGTCCCAACAGATTTATCCTCATGATTTATTTATCGACTTT 1020
1021 CTATGGTTGACAAATTTGAAAGATTTAGAAATTTCAACCAATTTCAAGAGTGGAAATTA 1080
1021 CTATGGTTGACAAATTTGAAAGATTTAGAAATTTCAACCAATTTCAAGAGTGGAAATTA 1080
1081 ATGTTTTAGATGAAACATACAGATGTTGGTGGACGAGATGAGCAAAATTTATCATGATG 1140
1081 ATGTTTTAGATGAAACATACAGATGTTGGTGGACGAGATGAGCAAAATTTATCATGATG 1140
1141 TTTGTAACATGCTTTAGCTTTTGGTTTAAAGATCAATGGTAAATTTAGAGTTTCCCGAG 1200
1141 TTTGTAACATGCTTTAGCTTTTGGTTTAAAGATCAATGGTAAATTTAGAGTTTCCCGAG 1200
1201 ATCCATTTGGCTGAAATTTACTAATGAAATTTAGCTTTGAAAGACGAAATTTAGAGCTTT 1260
1201 ATCCATTTGGCTGAAATTTACTAATGAAATTTAGCTTTGAAAGACGAAATTTAGAGCTTT 1260
1261 CATATCATGCTGACATATATATATCAAGAGATTTATCTTCTGAAAAACAAATCTTGA 1320
1261 CATATCATGCTGACATATATATATCAAGAGATTTATCTTCTGAAAAACAAATCTTGA 1320
1321 AGTCAGCTGATTTCTTCAAGAGATTAATATCACTGATTTCAACAGGCTTTCTAAATTA 1380
1321 AGTCAGCTGATTTCTTCAAGAGATTAATATCACTGATTTCAACAGGCTTTCTAAATTA 1380
1381 TTCAACAAGAGTGGAAAAATGCTTTAGTTTCCCTATCAATACCGTTTGAAGCGATAA 1440
1381 TTCAACAAGAGTGGAAAAATGCTTTAGTTTCCCTATCAATACCGTTTGAAGCGATAA 1440
1441 ACATGAGCAATATACAGCTTTTCAATGAGCAATATCAAGAAATTTCTGAAAACTACAT 1500
1441 ACATGAGCAATATACAGCTTTTCAATGAGCAATATCAAGAAATTTCTGAAAACTACAT 1500
1501 ATCACTCATCAAAATTTAGTAACTGATTTCACTGATTTGAGTGGTGGTAAAGATTTCTACA 1560
1501 ATCACTCATCAAAATTTAGTAACTGATTTCACTGATTTGAGTGGTGGTAAAGATTTCTACA 1560
1561 CTTGGCAATCTTATTTATCGTAAAGATTTAAAGTCTTGAAGTGGTGGTAAAGATTA 1620
1561 CTTGGCAATCTTATTTATCGTAAAGATTTAAAGTCTTGAAGTGGTGGTAAAGATTA 1620
1621 AGTTGGACCAAGCTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTGTTGCTG 1680
1621 AGTTGGACCAAGCTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTGTTGCTG 1680
1681 CAACACTTTCTGCTCCGAAATTTATCAGATGCGGTATTTATGCGGCAAAATTTGGCATAT 1740
1681 CAACACTTTCTGCTCCGAAATTTATCAGATGCGGTATTTATGCGGCAAAATTTGGCATAT 1740
1741 TAACACTAGTGTGATGATCTTTTTCATATCGTGTGATCAATCGATGAAATTTGACCAACC 1800
1741 TAACACTAGTGTGATGATCTTTTTCATATCGTGTGATCAATCGATGAAATTTGACCAACC 1800
1801 TGATTTCAATGTTGAAAAATTTGGAATTTAGATGTCGACAAAGATTTGTTGTTGAGAGATG 1860

| | | | |
|----|------|--|------|
| Db | 1801 | TGATTCAAATGTGTTGAAATAATGGAATGTAGATGTCGACAAGGATGTGTGTTACAGAGCATG | 1860 |
| Qy | 1861 | TTCCGAAATTTATTTTTCAGCATTAAGAAAGATGCAATCTGTGTGATTTGGAGATGAAAGCTTTTAA | 1920 |
| Db | 1861 | TTCCGAAATTTATTTTTCAGCAATTAAGAAAGATGCAATCTGTGTGATTTGGAGATGAAAGCTTTTAA | 1920 |
| Qy | 1921 | AATGCGAAGCCGGGAGTAACTAGCCAAATGTATTTCAAACTGTGTTGGATTTGGAACTAATGAATA | 1980 |
| Db | 1921 | AATGCGAAGCCGGGAGTAACTAGCCAAATGTATTTCAAACTGTGTTGGATTTGGAACTAATGAATA | 1980 |
| Qy | 1981 | GTATGTTGAGAGAAGCTATATGGAACAAGAGATGCTTATGTGCCAAACATTAATAATGAATATA | 2040 |
| Db | 1981 | GTATGTTGAGAGAAGCTATATGGAACAAGAGATGCTTATGTGCCAAACATTAATAATGAATATA | 2040 |
| Qy | 2041 | TGGAAAACGCTTACGTGTCATTTGCCATTTAGGCCCGGATTTCAAGCCCGCTTATTTACTTTTG | 2100 |
| Db | 2041 | TGGAAAACGCTTACGTGTCATTTGCCATTTAGGCCCGGATTTCAAGCCCGCTTATTTACTTTTG | 2100 |
| Qy | 2101 | TGGGGGCCAAATTTATCAGAGAGATTTGTTGAAAGCTCTGAATATCATPAATCTATTTAAAGC | 2160 |
| Db | 2101 | TGGGGGCCAAATTTATCAGAGAGATTTGTTGAAAGCTCTGAATATCATPAATCTATTTAAAGC | 2160 |
| Qy | 2161 | TAATGACGACGACGGGTGCATCTCTAAACGATATCCATAGCTTCAAGAGGGAAATTTAAAGG | 2220 |
| Db | 2161 | TAATGACGACGACGGGTGCATCTCTAAACGATATCCATAGCTTCAAGAGGGAAATTTAAAGG | 2220 |
| Qy | 2221 | AAGCCAAATTTAAACCGCGGTAGCATTTGCATTTGAGTAAACGAGAGAAAGTGGGAAAGTGGAAAG | 2280 |
| Db | 2221 | AAGCCAAATTTAAACCGCGGTAGCATTTGCATTTGAGTAAACGAGAGAAAGTGGGAAAGTGGAAAG | 2280 |
| Qy | 2281 | AAGAGGTTGTGGAGGAGATGATGATGATGATTAATAAACAAGAGAGAAAGATTAATGAAAT | 2340 |
| Db | 2281 | AAGAGGTTGTGGAGGAGATGATGATGATGATTAATAAACAAGAGAGAAAGATTAATGAAAT | 2340 |
| Qy | 2341 | TAATTTTGTGAAGAAAAATGTTAGCATTTCTCTAGAGCTTCTTAAAGATGCAATTTTGGAAAC | 2400 |
| Db | 2341 | TAATTTTGTGAAGAAAAATGTTAGCATTTCTCTAGAGCTTCTTAAAGATGCAATTTTGGAAAC | 2400 |
| Qy | 2401 | TGTGTACAGTGTGTAATTTTTTTTTTACGAAACGATGACCGGTTTACTTGGAAAACACGATTC | 2460 |
| Db | 2401 | TGTGTACAGTGTGTAATTTTTTTTTTACGAAACGATGACCGGTTTACTTGGAAAACACGATTC | 2460 |
| Qy | 2461 | TTGATACGTGTGAAGGACATCAATTTTAAACCCGTTGGTCTTGTGAATGAAAATGAAGAAC | 2520 |
| Db | 2461 | TTGATACGTGTGAAGGACATCAATTTTAAACCCGTTGGTCTTGTGAATGAAAATGAAGAAC | 2520 |
| Qy | 2521 | AAAGGTAATTTGGCCTTTTAAAGATGATAATATCATGCTCTTGAACGGGGTACTGTTTG | 2580 |
| Db | 2521 | AAAGGTAATTTGGCCTTTTAAAGATGATAATATCATGCTCTTGAACGGGGTACTGTTTG | 2580 |
| Qy | 2581 | TAGTTGTATATAAAGGTTGTAGTTGTATATAAAGGTAATAGGTAATCAATAGAAAGCT | 2640 |
| Db | 2581 | TAGTTGTATATAAAGGTTGTAGTTGTATATAAAGGTAATAGGTAATCAATAGAAAGCT | 2640 |
| Qy | 2641 | TAAGTTATTAAGTTTTTCCCTCGTGTGCACACCGGTGAGGTTCTGTTTAAAGCAGTTTAT | 2700 |
| Db | 2641 | TAAGTTATTAAGTTTTTCCCTCGTGTGCACACCGGTGAGGTTCTGTTTAAAGCAGTTTAT | 2700 |
| Qy | 2701 | TTTATGTTTAGTACGACACTTGGTGGTGTGTTGTTGTTTATTTCTCTAAAGAGGCTTTTCACT | 2760 |
| Db | 2701 | TTTATGTTTAGTACGACACTTGGTGGTGTGTTGTTGTTTATTTCTCTAAAGAGGCTTTTCACT | 2760 |
| Qy | 2761 | TTGTTCTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA | 2792 |
| Db | 2761 | TTGTTCTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA | 2792 |

RESULT 2
ADM98942
ID ADM
XX
AC ADM
XX

| | | |
|----|-------------|---|
| DT | 01-JUL-2004 | (first entry) |
| XX | | |
| DE | | Diterpene synthase DNA #2. |
| XX | | |
| KW | | Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene; |
| KW | | diterpene precursor; diterpene synthase; defence toxin; |
| XX | | volatile defensive signal; pollinator attractant; photoprotectant. |
| OS | | |
| XX | | Stevia rebaudiana. |
| FN | | US2004072323-A1. |
| XX | | |
| PD | | 15-APR-2004. |
| XX | | |
| PF | | 07-JAN-2002; 2002US-00041018. |
| XX | | |
| PR | | 05-JAN-2001; 2001US-0259880P. |
| XX | | |
| PA | | (MATS/) MATSUDA S P T. |
| PA | | (HART/) HART E A. |
| XX | | |
| PI | | Matsuda SPT, Hart EA; |
| XX | | |
| DR | | WPI; 2004-373921/35. |
| XX | | |
| PT | | New unicellular organisms comprising exogenous nucleic acids encoding a |
| PT | | geranylgeranyl pyrophosphate and a diterpene synthase, useful for |
| XX | | producing diterpenes and diterpene precursors. |
| XX | | |
| PS | | Claim 3; SEQ ID NO 362; 38pp; English. |
| XX | | |
| CC | | The invention relates to a unicellular organism for producing a diterpene |
| CC | | or diterpene precursor comprising an exogenous nucleic acid sequence |
| CC | | encoding a geranylgeranyl pyrophosphate synthase under the control of a |
| CC | | promoter operable in the organism, and an exogenous nucleic acid sequence |
| CC | | encoding a diterpene synthase under the control of a promoter operable in |
| CC | | the organism. The invention also relates to methods of producing a |
| CC | | diterpene or diterpene precursor and a method of isolating a diterpene |
| CC | | synthase comprising growing several cells in the presence of a |
| CC | | polyaromatic resin to make a cell/resin mixture, where at least one of |
| CC | | the cells further comprises at least one isolated and purified nucleic |
| CC | | acid sequence of a yeast expression library, and the expression of the |
| CC | | nucleic acid sequence is regulated by an inducible promoter under |
| CC | | conditions where the expression is induced, filtering the cell/resin |
| CC | | mixture, extracting the cell/resin mixture with alcohol to produce an |
| CC | | organic eluent and analysing the organic eluent by a screening method |
| CC | | including chromatography and/or spectroscopy, to identify the nucleic |
| CC | | acid sequence encoding the diterpene synthase. The unicellular |
| CC | | microorganism is useful as a diterpene or diterpene precursor producing |
| CC | | system. Diterpenes, in plants, serve as defence toxins, volatile |
| CC | | defensive signals, pollinator attractants and photoprotectants. This |
| CC | | sequence represents DNA encoding a diterpene synthase polypeptide of the |
| CC | | invention. Note: The sequence data for this patent did not form part of |
| CC | | the printed specification but was obtained in electronic format from |
| CC | | USPTO at seqdata.uspto.gov/sequence.html . |
| XX | | |
| SQ | | Sequence 3117 BP; 986 A; 502 C; 613 G; 1016 T; 0 U; 0 Other; |

| | Query Match | 92.5%; | Score 2582.6; | DB 12; | Length 3117; |
|----|-----------------------|--|---------------|----------------|--------------|
| | Best Local Similarity | 98.7%; | Pred. No. 0; | | |
| | Matches 2603; | Conservative | 0; | Mismatches 34; | Indels 0; |
| | | | | | Gaps 0 |
| QY | 140 | TTTGCATTTAGTAAATCAGCAAACTAAACATCAATGAATCTTTTCATCTATGCATCGGCTC | 199 | | |
| | | | | | |
| Db | 466 | TTGGTGTGCAGTAAATCAGCAAACTAAACATCAATGAATCTTTTCATCTATGCATCGGCTC | 525 | | |
| | | | | | |
| QY | 200 | CCCTTTGTAAACCAAAATCAAAATCGACCGCGGGCTCTGTCACTATTTCATACAGCATCAAC | 259 | | |
| | | | | | |
| Db | 526 | CCCTTTGTAAACCAAAATCAAAATCGACCGCGGGCTCTGTCACTATTTCATACAGCATCAAC | 585 | | |
| | | | | | |
| QY | 260 | TTCCATCTGTGTGGACAAACTAATCCCACTATCTTGATCATACACCAAAAGAACGGAT | 319 | | |
| | | | | | |
| Db | 586 | TTCCATCTGTGTGGACAAACTAATCCCACTATCTTGATCATACACCAAAAGAACGGAT | 645 | | |
| | | | | | |

| | | | | | | | |
|----|------|---|------|----|------|---|------|
| QY | 320 | CCAAAAACAGTTTAAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC | 379 | QY | 1400 | TGCTCTTAAGTTCCTATCAATACCGGTTTAGAAGCGATAAACACTAGACGAATATACA | 1459 |
| Db | 646 | CCAAAACTGTTTAAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC | 705 | Db | 1736 | TGCTCTTAAGTTCCTATCAATACCGGTTTAGAAGCGATAAACACTAGACGAATATACA | 1785 |
| QY | 380 | CATGTCCTCTCCAAACTCACCCAAATCGCTTGTTCCTGAGTGTCTCAATTTGTT | 439 | QY | 1460 | GCTTTACAACTGAGCAATACAAAGAAATCTCGAAACTACATATCACTCATCAAAATATTAG | 1519 |
| Db | 706 | CATGTCCTCTCCAAACTCACCCAAATCGCTTGTTCCTGAGTGTCTCAATTTGTT | 765 | Db | 1786 | GCTTTACAACTGAGCAATACAAAGAAATCTCGAAACTACATATCACTCATCAAAATATTAG | 1845 |
| QY | 440 | AATTAATCAATCAGTTAATGATGGTTCATGGGGTCTTGTTAATACACATCATATAATAA | 499 | QY | 1520 | TAACTCATTAATCACTAAGGTTGGCTTGTGAAGATTTCTACCTGCCAATCTATTATTGCG | 1579 |
| Db | 766 | AATTAATCAATCAGTTAATGATGGTTCATGGGGTCTTGTTAATACACATCATATAATAA | 825 | Db | 1846 | TAACTCATTAATCACTAAGGTTGGCTTGTGAAGATTTCTACCTGCCAATCTATTATTGCG | 1905 |
| QY | 500 | TCACCGTTCCTTAAAGATTCCTATCTTCAACATGACATGATTTGTCATTTAAAAAG | 559 | QY | 1580 | TGAAGAATTAAGGCTTGTGAAGAGTGGGTGTAGAGATAAGTTGGACCACTCAAGTT | 1639 |
| Db | 826 | TCACCGTTCCTTAAAGATTCCTATCTTCAACATGACATGATTTGTCATTTAAAAAG | 885 | Db | 1906 | TGAAGAATTAAGGCTTGTGAAGAGTGGGTGTAGAGATAAGTTGGACCACTCAAGTT | 1965 |
| QY | 560 | ATGGAATGTTGGGGAAGATCAAAATAAATAAGGCTTAAGTTCCTTATTTAGTCAAACTTGC | 619 | QY | 1640 | TGCTAGGCAAAAGACCGCTACTGTTATTTCTGCTGCTCAACACATTTTCCTCTCCCCA | 1699 |
| Db | 886 | ATGGAATGTTGGGGAAGATCAAAATAAATAAGGCTTAAGTTCCTTATTTAGTCAAACTTGC | 945 | Db | 1966 | TGCTAGGCAAAAGACCGCTACTGTTATTTCTGCTGCTCAACACATTTTCCTCTCCCCA | 2025 |
| QY | 620 | TTTCACTACTGAAAAAAGTCAACCAATCTCCATTTGGTTTGGATCATATTTTCTGGTTT | 679 | QY | 1700 | ATTATCAGATGCGCGTATTTTCATGCGGCCAAAAATGGCAATTAATTAACATGATGTTGATGA | 1759 |
| Db | 946 | TTTCACTACTGAAAAAAGTCAACCAATCTCCATTTGGTTTGGATCATATTTTCTGGTTT | 1005 | Db | 2026 | ATTATCAGATGCGCGTATTTTCATGCGGCCAAAAATGGCAATTAATTAACATGATGTTGATGA | 2085 |
| QY | 680 | GCTTGAGTATGCGAAAACTTGGACATAAACCCTCTTTCAAAAACAAACAGATTTTAGTTT | 739 | QY | 1760 | CTTTTGTGATATCGGTGGTGTAGAAATCGATGAAATGACCACTGATTCATATGTTGAAAA | 1819 |
| Db | 1006 | GCTTGAGTATGCGAAAACTTGGACATAAACCCTCTTTCAAAAACAAACAGATTTTAGTTT | 1065 | Db | 2086 | CTTTTGTGATATCGGTGGTGTAGAAATCGATGAAATGACCACTGATTCATATGTTGAAAA | 2145 |
| QY | 740 | GATGCTACATAAGAGGAATTTGAGCAAAAAAGATGCCATTCAAATGAGATGATGGATA | 799 | QY | 1820 | ATGGAATGATGATGTCGCAAGGATTTGTTGTCAGAGCATGTTTCGATTTATTTTATG | 1879 |
| Db | 1066 | GATGCTACATAAGAGGAATTTGAGCAAAAAAGATGCCATTCAAATGAGATGATGGATA | 1125 | Db | 2146 | ATGGAATGATGATGTCGCAAGGATTTGTTGTCAGAGCATGTTTCGATTTATTTTATG | 2205 |
| QY | 800 | CTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGTTGAAGAAATA | 859 | QY | 1880 | ATTTAAAGATGCAATCTGTTGGATTGGAGATGAAGCTTTTAAATGGCAAGCGCGCATGT | 1939 |
| Db | 1126 | CTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGTTGAAGAAATA | 1185 | Db | 2206 | ATTTAAAGATGCAATCTGTTGGATTGGAGATGAAGCTTTTAAATGGCAAGCGCGCATGT | 2265 |
| QY | 860 | TCAGATGAANAATGTTCTGTTTCAACTCACCATCAGCAACAGCTGCTTTCATTTAA | 919 | QY | 1940 | AACTAGCCATGTTTATTCAAACTTGGTGTGAACCTTAATGAATAGTATGTCGAGAGCACTAT | 1999 |
| Db | 1186 | TCAGATGAANAATGTTCTGTTTCAACTCACCATCAGCAACAGCTGCTTTCATTTAA | 1245 | Db | 2266 | AACTAGCCATGTTTATTCAAACTTGGTGTGAACCTTAATGAATAGTATGTCGAGAGCACTAT | 2325 |
| QY | 920 | TCAATCAAAATCTCTGTTGCTTAAATTTATTTAAATTCATTTTGGCAAGTTTGGTAATGC | 979 | QY | 2000 | ATGGAACAAGATGCTTTATGTCGCAACATTTAAATGAATATATGCAAAAACGCTTACGTC | 2059 |
| Db | 1246 | TCAATCAAAATCTCTGTTGCTTAAATTTATTTAAATTCATTTTGGCAAGTTTGGTAATGC | 1305 | Db | 2326 | ATGGAACAAGATGCTTTATGTCGCAACATTTAAATGAATATATGCAAAAACGCTTACGTC | 2385 |
| QY | 980 | AGTCCCAACAGTTTATTCCTCATGATTTATTTACCGACTTTCTATGTTGACACAAATGA | 1039 | QY | 2060 | ATTTGCAATTAGGCCCGCATTTCAAGCGCGCTATTTACTTTTGTGGGCGCCAAATATCAGA | 2119 |
| Db | 1306 | AGTCCCAACAGTTTATTCCTCATGATTTATTTACCGACTTTCTATGTTGACACAAATGA | 1365 | Db | 2386 | ATTTGCAATTAGGCCCGCATTTCAAGCGCGCTATTTACTTTTGTGGGCGCCAAATATCAGA | 2445 |
| QY | 1040 | AGATTTAGGAATTTTCAACCAATTTTCAAGTGGGAAATTTAAAAATGTTTATGATCAAAATGA | 1099 | QY | 2120 | GGAGATGTTTGAAGCTCTGAATATCATTAATCTTAATTTAAGCTAATGAGCACGAGGTCG | 2179 |
| Db | 1366 | AGATTTAGGAATTTTCAACCAATTTTCAAGTGGGAAATTTAAAAATGTTTATGATCAAAATGA | 1425 | Db | 2446 | GGAGATGTTTGAAGCTCTGAATATCATTAATCTTAATTTAAGCTAATGAGCACGAGGTCG | 2505 |
| QY | 1100 | CAGATGTTGGGTGGAGACGATGAGCAAAATATTCAATGATGTTGTAAATGCTGCTTATG | 1159 | QY | 2180 | ACTTTAAACGATATCCATAGCTTCAAGGGAAATTTAAGGAAGGCAAAATTAACCGCGGT | 2239 |
| Db | 1426 | CAGATGTTGGGTGGAGACGATGAGCAAAATATTCAATGATGTTGTAAATGCTGCTTATG | 1485 | Db | 2506 | ACTTTAAACGATATCCATAGCTTCAAGGGAAATTTAAGGAAGGCAAAATTAACCGCGGT | 2565 |
| QY | 1160 | CTTTCCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTC | 1219 | QY | 2240 | AGCATTTGCAATTTGAGTAAACGAGCAAAAGTGGGAAAGTGGAAAGAGAGTTGTGGAGGAGAT | 2299 |
| Db | 1486 | CTTTCCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTC | 1545 | Db | 2566 | AGCATTTGCAATTTGAGTAAACGAGCAAAAGTGGGAAAGTGGAAAGAGAGTTGTGGAGGAGAT | 2625 |
| QY | 1220 | TAATGAATTAAGCTTTGAAAGACCAATATGAGCTTCTTGAACATATCATGCTCACATAT | 1279 | QY | 2300 | GATGATGATGATTAATAAACHAAGAGGAAGAAATTAATGAATTAATTTTGAAGAAATGG | 2359 |
| Db | 1546 | TAATGAATTAAGCTTTGAAAGACCAATATGAGCTTCTTGAACATATCATGCTCACATAT | 1605 | Db | 2626 | GATGATGATGATTAATAAACHAAGAGGAAGAAATTAATGAATTAATTTTGAAGAAATGG | 2685 |
| QY | 1280 | ATTATACCAAGAGGATTTATCTCTGAAAAACAAATCTTGAAGTCAGCTGATTTTCTCAA | 1339 | QY | 2360 | TAGCATTTGTTTCTTAGAGCTTGTAAAGATGATGATTTTGAACATGTCACGCTTGAATTT | 2419 |
| Db | 1606 | ATTATACCAAGAGGATTTATCTCTGAAAAACAAATCTTGAAGTCAGCTGATTTTCTCAA | 1665 | Db | 2686 | TAGCATTTGTTTCTTAGAGCTTGTAAAGATGATGATTTTGAACATGTCACGCTTGAATTT | 2745 |
| QY | 1340 | AGAGATAATATCCACTGATTAACAGGCTTTCTTAAATTAATTCACAAAGAGGTGAAAA | 1399 | QY | 2420 | TTTTTACGCAACGATGACGGTTTACTGGAACAACGATTTCTTGATCTGTGAAGGACAT | 2479 |
| Db | 1666 | AGGATATATATCCACTGATTAACAGGCTTTCTTAAATTAATTCACAAAGAGGTGAAAA | 1725 | Db | 2746 | TTTTTACGCAACGATGACGGTTTACTGGAACAACGATTTCTTGATCTGTGAAGGACAT | 2805 |
| | | | | QY | 2480 | CAITTAACAACCGTTGCTGCTGTGAATGAAATGAAGAACAAAGGTAATTTGGCCTTTAA | 2539 |

Db 2806 CATTTACAAACCGTGTGCTGTGTAATGAAATGAAGAACAAAGTAATTGSCCTTTAA 2865
 Qy 2540 AAGATGATAAATATCATCTCTTCACGGGGTACTGTGTAGTTGTATATAAAGGTT 2599
 Db 2866 AAGATGATAAATATCATCTCTTCACGGGGTACTGTGTAGTTGTATATAAAGGTT 2925
 Qy 2600 GTAGTTGTATATAAAGGTAATAGTAATCAATAGAACCTTAAGTTATTAAGTTTTC 2659
 Db 2926 GTAGTTGTATATAAAGGTAATAGTAATCAATAGAACCTTAAGTTATTAAGTTTTC 2985
 Qy 2660 CTCGTGTACACACCGTGAGGTTCTGTGTAAAGCAGTTTATTTATTTAGTACGACAC 2719
 Db 2986 CTCGTGTACACACCGTGAGGTTCTGTGTAAAGCAGTTTATTTATTTAGTACGACAC 3045
 Qy 2720 TTGGTGGTGTGCTGTGTTATCTCTTAAGAGCCTTCACTTTGTTCTTAAAAAAA 2776
 Db 3046 TTGGTGGTGTGCTGTGTTATCTCTTAAGAGGCTTTCACCTTTGTTCTTAAATAGAA 3102

RESULT 3
 ADM98948
 XX ADM98948 standard; DNA; 2594 BP.
 AC
 XX
 XX ADM98948;
 DT 01-JUL-2004 (first entry)
 XX
 DE Diterpene synthase DNA #8.
 XX
 KW Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
 KW diterpene precursor; diterpene synthase; defence toxin;
 KW volatile defensive signal; pollinator attractant; photoprotectant.
 XX
 OS Lactuca sativa.
 XX
 FN US2004072323-A1.
 XX
 XX 15-APR-2004.
 XX
 XX 07-JAN-2002; 2002US-00041018.
 XX
 XX 05-JAN-2001; 2001US-0259880P.
 XX
 XX (MATS/) MATSUDA S P T.
 XX (HART/) HART E A.
 XX
 XX Matsuda SPT, Hart EA;
 XX
 XX WPI; 2004-373921/35.
 XX
 XX New unicellular organisms comprising exogenous nucleic acids encoding a
 XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for
 XX producing diterpenes and diterpene precursors.
 XX
 XX Claim 3; SEQ ID NO 368; 38pp; English.
 XX
 XX The invention relates to a unicellular organism for producing a diterpene
 XX or diterpene precursor comprising an exogenous nucleic acid sequence
 XX encoding a geranylgeranyl pyrophosphate synthase under the control of a
 XX promoter operable in the organism, and an exogenous nucleic acid sequence
 XX encoding a diterpene synthase under the control of a promoter operable in
 XX the organism. The invention also relates to methods of producing a
 XX diterpene or diterpene precursor and a method of isolating a diterpene
 XX synthase comprising growing several cells in the presence of a
 XX polyaromatic resin to make a cell/resin mixture, where at least one of
 XX the cells further comprises at least one isolated and purified nucleic
 XX acid sequence of a yeast expression library, and the expression of the
 XX nucleic acid sequence is regulated by an inducible promoter under
 XX conditions where the expression is induced, filtering the cell/resin
 XX mixture, extracting the cell/resin mixture with alcohol to produce an
 XX organic eluent and analysing the organic eluent by a screening method
 XX including chromatography and/or spectroscopy, to identify the nucleic

CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents DNA encoding a diterpene synthase polypeptide of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification but was obtained in electronic format from
 CC USPTO at seqdata.uspto.gov/sequence.html.
 XX

Qy Sequence 2594 BP; 812 A; 444 C; 532 G; 806 T; 0 U; 0 Other;
 Query Match 51.2%; Score 1428.6; DB 12; Length 2594;
 Best Local Similarity 75.8%; Pred. No. 0;
 Matches 1833; Conservative 0; Mismatches 529; Indels 57; Gaps 3;
 Qy 238 CAGCTATTTCATACAGCAATCACTTCATCATGTGTGACAACTAAATCCCACTAATCTGATCA 297
 Db 187 CATCGGGTCTTCGTACAGCTTCTTCACAAGCTGGACAAAGTTAATCCTACTGTCATGACCC 246
 Qy 298 TTGATACAAACCAAGACGGATCCAAAAACAGTTTAAAAATGTAGAAATTTCTGTTCTT 357
 Db 247 TTGATGTGACCAAAAGACGAATCCGAAAAGCTGTTCAACAATGTGGAAGTTTCTGTTCTT 306
 Qy 358 CATATGACACAGCATGGGTAGCCATGGTCCCTTCTCCAAACTCACCCTAAATCGCTTGT 417
 Db 307 CATATGACACAGCTTGGGTAGCCATGGTCCCTTCTCCAAACTCACCCTAAATCGCTTGT 366
 Qy 418 TCCCTGAGTGTCTCAATGGTTAATTAATATCAAGCTTAAATGATGTTTCATGGGCTCTTG 477
 Db 367 TCCCTGATGTCTGAACCTGGTTACTGGATAATCAAGCTTGAATGATGTTTCATGGGCTCT-- 424
 Qy 478 TTAATCACACTCAATATCATATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 537
 Db 425 -----TCTTCTCTCATCAGTCCCAATTAATTAAGATACTCTCTCTTCAACATTAG 474
 Qy 538 CATGTATTGTTGCAATTAATAAGATGGAATGTTGGGGAAGATCAATAATAATAAGTCTAA 597
 Db 475 CATGTGTACTTGCATTAATAAGATGGAATGTTGGGGAAGATCAATAATAATAAGTCTAA 534
 Qy 598 GTTTTATTTGAGTCAAAATCTTCTTCTCAGCTACTGAAAAAAGTCAACCACTCTCCCAATG 657
 Db 535 ATTACATTTAGTCAAAATTTTCTTCTCAGTCACTGACAAAAATCAAGCACTCTCCATTTG 594
 Qy 658 TTGACATCATATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 717
 Db 595 TTGACATCATATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 654
 Qy 718 CAAAAACAAACAGATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 777
 Db 655 TAAACCAACACATTTGAGTGTGATGTTCAATGAGAGAGATTTGAGCTAAGAGATGTC 714
 Qy 778 ATTCAAAATGAGAGATGAGTATCTGCGTATATCTCTGAGGAGCTCGGTAATTTATATG 837
 Db 715 ATTCAAATGGGAGGAGCATCTTGGCATATATCTCAGAGGAGCTTGGAAATTTAAATG 774
 Qy 838 ATTGGAATATGTTGCAACAAATATCAGATGAAATGTTCTGTTTCAACTCACCATCAG 897
 Db 775 ACTGGAACATGGTGTGATGAAATATCAAAATGAAGATGTTCTCTTTTCAACTCACCCTCAG 834
 Qy 898 CAACAGCTGCTGCTTTCATTAATCATCAAAATCTCTGTTGTTGTTGTTGTTGTTGTTGTT 957
 Db 835 CAACAGCTGCTGCTTTCATTAATCATCAAAATGCTGTTGTTGTTGTTGTTGTTGTTGTT 894
 Qy 958 TTTTGGCAAGTTTGGTAAATGCAAGTTCCTCAAGCTTTATCTCATGATTTATTTATCCGAC 1017
 Db 895 TCTTTGGCAAAATTTGGAATGCAAGTTCCTCAAGCTTTATCTCATGATTTATTTATCCGAC 954
 Qy 1018 TTCTATGTTGACACAAATTTGGAATGCAAGTTCCTCAAGCTTTATCTCATGATTTATTTATCCGAC 1077
 Db 955 TTTTCCATGTTGACACACTTTGGAATGCAAGTTCCTCAAGCTTTATCTCATGATTTATTTATCCGAC 1014
 Qy 1078 AAAATGTTTATGATGAAACATACAGATGTTGGGTGGAGACGAGATGACCAATATTCATGG 1137

1015 AAAATGTTTTAGATGAACATATAGATGTTGGGTGCAGGGGATGTGCAAAATATTCATGG 1074
1138 ATGTTGTAACATGTGCTTTAGCCTTTTCGTTATTAAGATCAATCGGTATGAAGTTTCCC 1197
1075 ATGTTGTACATTTGTGCTTTTAGCCTTTTCGCGTATTAAGAGCAATCGGTATGAAGTCTCTT 1134
1198 CAGATCCATTTGGCTGAAATTTACTAA-----TGAATTTAGCTTT 1233
1135 CGGATCCATTTGGCTAAAATTCACAAAGAGGGGATTAATGATGATTTCACTCGAAAGCCTT 1194
1234 TGAAGAGCAATATGATGAGCTCTTGAACAATATCATGCGTCAATATATATATATATCAAGAGG 1293
1195 TCAAGAGCGTATACACATCTCTTGAAGTTTATAAGGCATCAAGATCATATATCAAGAGG 1254
1294 ATTTATCTTCTGAAAAAACHAATCTTGAAGTCAAGTCAAGTCTTCTCAAGAGATAATATCCA 1353
1255 AGTTAGCTTTTAGAGAACAAAACCTTGA-----CAAGTTATCTCC 1293
1354 CTGATTTCAACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAATATGCTCTTAAGTTCC 1413
1294 CTTCAAGTAAACAGCTCTCTAATATATATCTCAAGAGGTGATGATGCTCTTAAGTTCC 1353
1414 CTATCAATACCGTTTGAACCGCATAAACATAGACGAATATACAGCTTTTCAATGTAG 1473
1354 CTTTTAATGGCAGCTAGAACGATGTCCACTAGAGAAACATAGAGCATTTACAATTTAA 1413
1474 ACAATACAGAAATCTGAAACCTTACATATACATCAATATATAGTAACTATGACATGATACC 1533
1414 ACCATACAGAAATCTTGAACCTTACATATAGCTCAACAACTATAGCAACAAAGATTACC 1473
1534 TAAAGTTGGCTGTGAAGATTTCTACACCTGCCAATCTATTTATCTGGAAGAAATTTAAAG 1593
1474 TAAAGTTAGCTGTGCAAGATTTCAATGAATGCCAATCTATCTATTTGGAAGAACTAAAG 1533
1594 GTCTTGAAGAGTGGGTGGTAGAATAAGTTGACCAAGCTCAAGTTTGTAGGCAAAAGA 1653
1534 ATCTTGAAGAGTGGGTGGTAGAACAAGATTAGACAAAGCTCAAAATTTGTAGACAAAGA 1593
1654 CGGCTTACTGTTATTTCTGTTGCTGCAACATTTGCTCCGCAATTTATCATGATGCGC 1713
1594 CAGCCTTACTGTTATTTCTGCTGCTCAATTTCTTCTCCCTGATTTATCATGATGCCC 1653
1714 GTATTTTCATGGGCAAAAATGCGATATTAACATGATGATGATGATGATGATGATGATG 1773
1654 GAAATTCATGGGCAAAAGTAGATCTCTACATGATGATGATGATGATGATGATGATG 1713
1774 GTGGTACAATCGATGAATGACCAACCTGATTCATGTTGTTGAAAATGGAATGATGATG 1833
1714 GAGGATCTATGGATGAATGCTGTTGAACTTTTGTTCACATCATTTGAAATAATGGAATGTAACG 1773
1834 TCGACAAAGATTTGTTTCAGACATGTTCCGATTTTATTTTATGATTAAGATTAAGATGCAA 1893
1774 TTGAAATGATTTGCTGTTTCAGAGAGATTTGGCGTTTATTTTATGACATTTAAGATGCGAG 1833
1894 TCTGTTGGATTTGAGATGAAGCTTTTAAATGGCAAGCGCGCTGATGTAACCTAGCCATGTTA 1953
1834 TCTGTTGGATTTGAGACAAAGCTTTTAAAGATTCAAGAACGCAATATAACCAAGCCATGTTA 1893
1954 TTCAAACTTGGTTGGAACTAATGAATGATGATGATGATGATGATGATGATGATGATGATG 2013
1894 TTGAAATTTGGTTGGATTTGGTGAAGAGTATGTTGAGAGAGCTATATGCGGCAAGATG 1953
2014 CTATGTTGCCAATTAATGAATATATGAAACGCTTACGTGTCATTTGCTATTTAGGCCC 2073
1954 GTTCAATTTCCAAATAAATGAATATATGAAATGTTTACGTGTCATTTGCTTTAGGCC 2013
2074 CCAATGTCAGCGGCTTATTTACTTTGTTGGGCCCCAAATTTATCAGAGAGATTTGTTGAAA 2133
2014 CGATGTCCTCCCTACTCTTTACTTTTATGAGGTTTAAATTTGTCGGAAGAGTTGTTTACA 2073
2134 GCTCTGAATATCAATCTATTTAAGCTATAGCAGCGAGGTCGATCTCTTAACAGATA 2193
2074 GCTCCGAGTATCAAGCTTTATGAGTTTATGAGCTTATGAGCACTCAGGTCGTTCTCATGAATGATA 2133

2194 TCCATAGCTTCAAGAGGAAATTTAAGGAAGCGCAAAATTAACCGGTAGCATTTGATTTGA 2253
2134 TCCATAGCTTCAAGAGGAAAGAGCGGGAAATTTGAATGCTGTGCGATTTGATCA 2193
2254 GTACCGGAGAAAGTGGGAAAGTGGGAAAGAGGTTGTGAGGAGATGATGATGATTA 2313
2194 GTGATGGAAGAGTGGGAGTGTGGAAGAGAGGTTGTAGAGGAGATGAAGATTTTGACAA 2253
2314 AAAACAAGAGGAAAGAAATTAATGAATTAATTTTGAAGAAATGTTAGCATTTGTTCCTA 2373
2254 AAAGTCAAGGAAAGAGATGATGAATTAGTTTGGAAACAAAGAGAGTGTGTTCCAA 2313
2374 GAGCTTGTAAAGATGCAATTTTGGAAACATGTGTCAAGTGTGTAATTTTTCGCAACG 2433
2314 GAGTATGCAAGATGATTTTGGAAACATGTGTCAAGTGTGTAATTTTTCGCAACG 2373
2434 ATGACGGGTTTACTGAAACACGATTTCTTGATCTGTGAGGACATCATTTCAACCCGT 2493
2374 ATGATGGGTTCACTGGAATGCCATTTCTTGATGTTGTCAAGGAAATATATATGAACCTG 2433
2494 TGGTCTGTGTAATGAATGAAGAAAGGTAATTTGGCCCTTTAAAGATGATAATAAT 2553
2434 TGTCCATGAGTTGATATTAATCATTTGTAACCAATTCGATATATATATATGATGATCTC 2493
2554 ATCATGCTCTTGCACGGGGTACTGTTGTAGTTGTATATAATAAGGTTGTAGTTGATATA 2613
2494 AAAAATGAACCAAGCATACTAGTATTTCTAGATCTACTCATGTTTATTTATTTTATTTT 2553
2614 AAGGTAATAGGTAATCAAT 2632
2554 AATATAGTTTGTAGCTAT 2572

RESULT 4

ADM98947
ID ADM98947 standard; DNA; 2638 BP.

AC ADM98947;

XX 01-JUL-2004 (first entry)

XX Diterpene synthase DNA #7.

XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;

KW diterpene precursor; diterpene synthase; defence toxin;

KW volatile defensive signal; pollinator attractant; photoprotectant.

XX Cucumis sativus.

XX US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX New unicellular organisms comprising exogenous nucleic acids encoding a

PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for

XX producing diterpenes and diterpene precursors.

PS Claim 3; SEQ ID NO 367; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene

CC or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
 CC promoter operable in the organism, and an exogenous nucleic acid sequence
 CC encoding a diterpene synthase under the control of a promoter operable in
 CC the organism. The invention also relates to methods of producing a
 CC diterpene or diterpene precursor and a method of isolating a diterpene
 CC synthase comprising growing several cells in the presence of a
 CC polyaromatic resin to make a cell/resin mixture, where at least one of
 CC the cells further comprises at least one isolated and purified nucleic
 CC acid sequence of a yeast expression library, and the expression of the
 CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents DNA encoding a diterpene synthase polypeptide of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification but was obtained in electronic format from
 CC USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 2638 BP; 810 A; 456 C; 581 G; 781 T; 0 U; 0 Other;

Query Match 29.0%; Score 809.4; DB 12; Length 2638;
 Best Local Similarity 62.0%; Pred. No. 1e-180;
 Matches 1410; Conservative 0; Mismatches 811; Indels 54; Gaps 6;

QY 267 GGTGACAACTAATCCCACTAATCTGATCATATGATACAAACCAAGACGGATCAAAAA 326
 Db 159 GGCACAAAACTAAACTCGGAGCTTTGGCTTTGAAGAAACAAAGAAAGAAATAAAAA 218
 QY 327 CAGTTTAAATATGAGAAATTTCTGTTCTTCATATGACACAGCATGGGTAGCCATGCTC 386
 Db 219 TTGTTCAAAATGTTGAACCTTTCAATTTTCGCATATGATCTGATGGGTGGCAATGGTC 278
 QY 387 CTTTCTCCAACTCAACCAATCGCTTGTCTGCTGAGTCTCTCAATTTGTTAATTAAT 446
 Db 279 CTTTCTCCAACTCTCTTAAACCTTTTCTGAGTGCATAACTGGGTATTTAGAT 338
 QY 447 AATCAGCTTAATGATGGTTCAATGGGTTCTGTTAATCACTCAATCAATCAATCAACCG 506
 Db 339 CATCAAAACCCGTGAGTCTGATGGG-----CATCTCCATGACCATCAGTTG 386
 QY 507 TTGCTTAAAGATTTCTATCTTCAACATAGCATGATTTGTCATTAAGAAAGATGAAT 566
 Db 387 GTGATGAAGCCACTCTCTTATCCACATAGCATGTGTCTTACTCTTAAGCCGATGGAT 446
 QY 567 GTTGGGGAAGATCAATAATAAAGGTTCTAAGTTTATTTAGTCAAAATCTTGTCTTCACT 626
 Db 447 ATCGTGTGATCATATGACGCAAGCCCTTAGTTTATCAAGTCTAATATAGCTTCACT 506
 QY 627 ACTGAAAAAGTCAACCATCTCCCATTTGGTTTGGATCATATTTCTGGTTTCTTGGAG 686
 Db 507 ACTGATGAGATCAACGTTCTCTGTGGGATTTGACATAATTTTCCCTGGTATGANTGAG 566
 QY 687 TATGCGAAAACTTGGACATAAACCCTCTTCAAAACAAACAGATTTTAGTTGATGCTTA 746
 Db 567 TAUGCTAAAGACTTGAATTTGAATCTACCTTTGGCATCAATGAATGGATGCTTTGGTT 626
 QY 747 CATAAGAGGAATTTGGAGCAAAAAAGAT-----GCCATTCAAATGAGATGGATGATAC 800
 Db 627 CAAAGAAAGAGTTGGAGCTTAGAGCTGCTGTAGCACTCTGAGAGGAAAGACCTAT 686
 QY 801 TTGCGGTATATCTGAGAGCTCGGTAATTTATATGATTTGGAATATGTTGGAAGAAATAT 860
 Db 687 TTACGGTATGTTTTCAGAGGAATTTGAAATTTACAGGACTGGGAAATGTTTCATGCGATAT 746
 QY 861 CAGATGAAATGTTCTGTTTCAACTCACCATCAGCAACAGCTGCTGCTTCAATTAAT 920
 Db 747 CAAAGGAAGACCGATCACTGTTTGTAGTTCTCCATCCACCGGCGATGGCTTTTATGCA 806

QY 921 CATCAAAATCCTGGTGTCTTAATTTAATTTCAATTTGGACAAGTTTGGTAATGCA 980
 Db 807 AGAAATGATGAGGCTGTTTAAATTTACCTTCGCTCAGTCTTACAAAAGTTTCATAGTTCA 866
 QY 981 GTCCCAACAGTTTATCCCTCATGATTTATTTATCCGACTTTCTATGTTTGACAAATTTGAA 1040
 Db 867 GTTCCCGCAATATATCTCTTGATATATATGCTGTTTACACATGTTGATGACCTTCAA 926
 QY 1041 AGATTAGAAATTCACACCATTTCCAGATGGAATAAATAATGTTTATAGATGAACATATAC 1100
 Db 927 AAACCTGGGATTTGATGGCCATTTCAAAGATGAGATTAGAGTGTATTAGATGAACATATAC 986
 QY 1101 AGATGTTGGGGAACGAGATGACCAATATTTCAATGATTTGATGATTTGATGCTTTAGCC 1160
 Db 987 AGCTGTTGGATGCAAGGGGAGGAAACATATTTCTTAGATGCTTCAACTTTGTGCAATGGCC 1046
 QY 1161 TTTCCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGGAATTAAT 1220
 Db 1047 TTCCGGATGTTACGTTTGAAGGATATGATGTTTCTTCAGATCAATTTGACTCAATTTCTCA 1106
 QY 1221 AA-----TGAAATGCTTTGAAAGACGAATATGACGCTTCTGAA 1259
 Db 1107 GAAGGTCTCTTTTCAAATTTGCTCGGAGGACATTTAAAGAGACTTTAGTGCCTCACTAGAG 1166
 QY 1260 ACATATCATGCTCACA-----TATATTATACCAAGAGGATTTATCTTCTGGAACAAATTC 1316
 Db 1167 TTATTTAAGGCTCTCCAGATTAATCAATTTATCCGGATGAGTTTATTTCTGGAATAATAAAC 1226
 QY 1317 TTGAAGTCAGCTGATTTTCTCAAAGAGATTAATATCCACTG-----ATTCAAAACAGG 1367
 Db 1227 TCTTGGACTAGTCTGTTTCTCGAATCATGATTAATCTAGTGTTCAGTTTCATTTCTGATAGA 1286
 QY 1368 CTTTCTTAAATTAATTCACAAAGAGGTGGAATAATGCTCTTAAAGTTCCCTTATCAATACCGGT 1427
 Db 1287 ACTCAGAGACTCGTGAACCAAGAGGAGTAAATGCTTTGAGTTTCCCTATATATTTCAACT 1346
 QY 1428 TTGAACGCGATAAACACTAGACGGAATATACAGCTTTTACATGTGTAGACAAATACAGATTT 1487
 Db 1347 CTGAAACGCTTATCAAAATTAAGCGAGCACTGGAAAGTTTACAGTGAGACATTTGTGAGGAT 1406
 QY 1488 CTGAAATCTACATATCACTCATCAATATTTAGTAACACTGATTAACCTTAAGTTTGGCTGTT 1547
 Db 1407 TCAAAACAGCATATGCTGCTTAAATTTTGGTTCATCAAGATTTCTTGGAACTTGTCTGTA 1466
 QY 1548 GAAGATTTTACACTGCTGCCAATCTATTTATGCTGAAGAAATTTAAAGGTCTTGAAGGTGG 1607
 Db 1467 GAAGATTTCAATACCTCTGCAAGGCATACATCGCAAGGAACTGAAAGAGCTTGAATAATGG 1526
 QY 1608 GTGCTAGAGAAATAGTTGGACCGACTCAAGTTTCTAGGCAAAAGACCGCTTACTGTTAT 1667
 Db 1527 GTCAATCAAAACAAATTTGGACAAGTTTGAATTTGGGAGACAGAAATTTAGCGTACTGCTAT 1586
 QY 1668 TTTCTGTTGCTGCAACACTTTGCTCTCCGAATTTATCAGATGCGCTATTTTCATGGGCC 1727
 Db 1587 TTTTCTGCTGACGAGACCTTAACTCTCCAGAACTTTTGTGATGCCCTTATCATGSGCA 1646
 QY 1728 AAAAATGGCAATTAATCTACATGATTTGATGACTTTTTTGTGATATCGGTGGTACAAATCGAT 1787
 Db 1647 AAAAATGGGTACTCAACACCGGTGTTGATGATTTCTTGTATGTTGAGGATCTGGAAGAG 1706
 QY 1788 GAATTCAGCAACCTGATTTCAATGTTTGAATAATGGAATCTAGATCTCGACAGGATTTGT 1847
 Db 1707 GAATTTGTAACCTTATACAAATTTGGTGAATAAGTGGGATGCCAGTGGGAAACCGGTATC 1766
 QY 1848 TGTTCAGAGCATGTTTGGGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGATTTGGA 1907
 Db 1767 TGTTCAGAGGAGTTGAGATTTATTTCTTGGCACTTCATAGCAATTTGTGAAATAGGA 1826
 QY 1908 GATGAAGCTTTTAAATGGCAAGCGCGATGTAACCTAGCCATGTTATTTCAAACTTGGTTG 1967
 Db 1827 AAAAAGCTTTTACCTTGGCAAGGACGCGGTGATGAGGAATGTTATCGATATTTGGTTG 1886
 QY 1968 GAATTAATGATGATGTTGAGAGAGCTATATGACAAAGATGCTTATTTGTGCCAACA 2027

1887 GCTTGTCTCGATCAATGAGGAGAACTGAATGTTGAAATAAGTAGTGCATCA 1946
2028 TTAATGAATATATGGAAGAACTTACGTGTCATTTGATTTAGGCCCGATTTGCAAGCG 2087
1947 TTGGATGAGTACATGAAATGGCTATGTATCATTTGCTTTGGGACCTATAGTCTTCCA 2006
2088 GCTATTTACTTTGTGGGCCCAATATATCAGAGGAGATTTGTTGAAAGCTCTCAATATCAT 2147
2007 AGCTCTTACTTTGTGGACCTAAGCTTCCAGAGGAAATTTGTTGAAATTTGAAATACAG 2066
2148 AATCTATTTAAGCTAATGAGCAGCAGGCTGCACTTCTAAACGATATCCATAGCTTCAAG 2207
2067 AGCTCTTTAGCTGATGACACTTCTGCGCCCTTCTGAATGATCTCGAATTTTGAT 2126
2208 AGGAAATTAAGAGAGGCAATTAATAACGCGGTAGCATTTGCATTTGAGTAAACGAGAAAT 2267
2127 AGAGAGTCCAGCGAAGGAAATTAATAATGCCCTTGTCTCTATACATGATTAGTCCCGTGT 2186
2268 GCGAAGTGGAAAGAGAGCTTGTGAGGAGATGATGATGATTAATAAACAAGAGGAAA 2327
2187 AAGCTCACCAAGAGAGGCCACTGAGCAATGAAAGAGATGTTGATAGAGAGAGAGA 2246
2328 GAATTAATGAATTAATTTTGAAGAAATGTTAGCATTTGTTCTTAGAGCTTTGTAAGAT 2387
2247 GAATGTTGAGATTAGTTTTCAGGAGAAACAGTACAA---TTCCAAGAGCTTGTAAAGAT 2303
2388 GCATTTGGAACATGTGTACAGTGTGATTTTTCAGCAACAGTACAGGCTTACT 2447
2304 TTGTTCTGGAATAGCTGTGTGTGCACTATTTTACAGGAAAGATGATGGGTTTACA 2363
2448 GGAACACAGATCTTGATCTGTGAGGACATCATTTACACCCGTTGGTGTG 2502
2364 TCCCATGAGTTGATGAATCTCTCGGAAAGCTTATTTGAACACCCATGTTCTGG 2418

RESULT 5
ADM98958
ID ADM98958 standard; DNA; 2658 BP.
XX ADM98958;
AC ADM98958;
XT 01-JUL-2004 (first entry)
DE Diterpene synthase DNA #18.
XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
XW diterpene precursor; diterpene synthase; defence toxin;
XW volatile defensive signal; pollinator attractant; photoprotectant.
XX Unidentified.
OS US2004072323-A1.
XX 15-APR-2004.
XX 07-JAN-2002; 2002US-00041018.
XX 05-JAN-2001; 2001US-0259880P.
XX (MATS/) MATSUDA S P T.
XX (HART/) HART E A.
XX Matsuda SPT, Hart EA;
XX WPI; 2004-373921/35.
XX New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX Claim 3; SEQ ID NO 378; 38pp; English.
XX

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents DNA encoding a diterpene synthase polypeptide of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2658 BP; 757 A; 486 C; 626 G; 789 T; 0 U; 0 Other;
Query Match 28.0%; Score 782.2; DB 12; Length 2658;
Best Local Similarity 61.3%; Pred. No. 2.7e-174;
Matches 1390; Conservative 0; Mismatches 823; Indels 54; Gaps 6;
XX
QY 275 AACTAATCCCACTAATCTGATCATTAACAACCAAGAACCGATCCAAACACAGTTTAA 334
DB 236 AACTAATACTGGAGCTTTGCACTTTGAAGAAACAAAGAAATTAAGAAATTTGTCGA 295
QY 335 AATGTAGAAATTTCTGTTTCTTATATGACACAGATGGGTAGCCATGCTCCCTCTCC 394
DB 296 CAAGGTTGAACCTTCAGTTTCTGCAATATGATGATGATGATGATGATGATGATGAT 355
QY 395 AACTCACCACAAATGCGCTTGTTCCTGAGTGTCTCAATTTGTTTAATTAATCAGCT 454
DB 356 AACTCTCTCAACCAACCTCTTTTCCCGAGTGTATTAATGGTATTAGATGATCAACA 415
QY 455 TAATGATGTTTCATGGGTCCTTGTGTTAATCACTCATATATATATATATATATATAT 514
DB 416 TCGTGTGCTCATGGGCTACTTCCACAACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 463
QY 515 AGATTTCTTCTTCTTCAACATTTAGCATGTTATTTGTCATTTAAAGATGAAATCTGGGA 574
DB 464 GGCCATCTCTTATCTACATTTAGCATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 523
QY 575 AGATCAATTAATAAAGGCTTAAGTTTATTTGAGTCAATCTTCTTCTTCTTCTTCTTCTTCT 634
DB 524 TGTATATATGAGCAAGGCCCTTGAATTTATCAAGTCTATATAGCTTCACTGATGA 583
QY 635 AAGTCAACCATCTCCCATTTGGTTTTCATATATTTCTGTTTCTGTTTCTGTTTCTGTTTCT 694
DB 584 GAACCAACGCTTCTCCGGTGGATTTGACATTTATTTTCCCTGGCATGATTTGAGTATCTTA 643
QY 695 AACTTTGGACATTAACCTCTTCTTCAACAAACACAGATTTTATTTGATGCTCATATAGAG 754
DB 644 AGACTTTGAATTTGAATCTACCTTGGCACCGAGAACGTTGGATGCTTGGTTTCCAAAGAA 703
QY 755 GGAATTTGGAGCAAAAA-----AGATGCCATTTCAAAATGAGATGGATGATGATGATGAT 808
DB 704 AGATTTGGAGCTCAGAGCTGCGAAGCAACTCTGAAGGTGGAAACCTTATTTAGCGTA 763
QY 809 TATCTCTGAAGCTCGGTAAATTTATGATTTGGAATATGCTGAAGAAATATCAGATGAA 868
DB 764 TGTTCAGAGGAAATTTGGAAGTTACAGGACTGGGATATGTTGTCATGATATCAAGGAA 823
QY 869 AAATGGTTCGTGTTTTCAACTCACCATCAGCAACAGCTGCTGCTTCTTCTTCTTCTTCTTCT 928

Db 824 GAATGATCACTGTTTAAATCTCCATCCACTAGCGCAGCGGCTTTTATGCATAGAAATGA 883
Qy 929 TCCTGCTGTCTTAATTAATTTAAATCACTTTTGACAGATTTGGTAAATGCGATCCCAAC 988
Db 884 TGAUGCTGTTTGAATATCTTCGCTCACTCTTACAAAGTTTGAATGGCTCAGTTCCCAAC 943
Qy 989 AGTTTATCCTCATGATTTATTTATCCGACTTTCTATGTTTGCACAAATGAAAGATTAGG 1048
Db 944 AATATATCCTCTTGATATATATGCTCGATTACACATGTTGATAGCCCTCAAAATTCGG 1003
Qy 1049 AATTTTCAACCACTTTTCAGAGTGAATTAATAATGTTTGAATGAACATACAGATGTTG 1108
Db 1004 AATTCCTGGCAATTTCAAGAGGAGATTAAGAGCGTATTAGATGAACCTTACAGGTGTTG 1063
Qy 1109 GGTGAACAGAGATGAGCAATATTTCAATGATGTTGTAACATGCTTTAGCCCTTTGGTT 1168
Db 1064 GATGCAAGGAGGAGAAATATATTTCTTAGATGCTTCACTTGTGCAATGGCCCTTTGGAAT 1123
Qy 1169 ATTAAAGGATCAATGGGTATGAATTTCCCGAGATCCATTTGGCTGAAATTAATAA 1222
Db 1124 GTTACGTTGTCAGGATATGATGTTTCTTCAGACCACTTGAATTTTCAAGATAT 1183
Qy 1223 -----TGAATTAGCTTTGAAAGAGATATATCCAGATCCATTTGGCTGAAATTAATAA 1267
Db 1184 CTTTCCCAATTCGCTTGAGGATATTTAAAGACTTCGGTGCCTCGCTGAGTTATATAA 1243
Qy 1268 TGCCTCACATATATATACCAAGAGGATTTATCTTCT---GGAACAACAACTTTGAAGTC 1324
Db 1244 GGCCTCTCAGATATACGACCCCGATGAATCTGTTCTGGAATAATAAATCTTTGGAC 1303
Qy 1325 AGCTGATTTCTCAAAGAGATATATATCCACTGAATTTCAAACAGGCTTTCTAAA 1376
Db 1304 TAGTCGTTTCTGAGCATGATATATCTAGTGAATTCAGTTTGGTCTGTAGAACGATAG 1363
Qy 1377 -TTAAATTCACAAAGGTGGAAATGCTTTAAGTTCCTATCAATACCGGTTTAGAACG 1435
Db 1364 TGTGTTTAAACAGAGGCTGTTAATGCTCTTGAATTCCTCCCTATATATGCAACTAGAACG 1423
Qy 1436 CATAAACACTAGACAAATATACAGCTTTTACAATGTAGACAAATCAAGAACTTCGAAAC 1495
Db 1424 CCTAATAGTAAGAGGCAATGGAAGTTACAGTGGAGCAATGTGAGGATTTCAAATC 1483
Qy 1496 TACATATCACTCATCAATATTAGTAACTGATTAACCTAAGTGGCTGTTGAAGATTT 1555
Db 1484 GCCATATGCTGCTTAAATTTGGCCATCAAGATTTTCTGGAACCTGCTGTAGAGGATTT 1543
Qy 1556 CTACACCTGCCAATCTATTTATCGTGAAGATTAAGAGTCTTGAAGTGGGTGGTGA 1615
Db 1544 CAATACCTGCAACGCAATCACTTAAAGAACTGGAAGAGCTTCAAAGATGGGTGGTTGA 1603
Qy 1616 GAATAGTTGGACCACTCAAGTTTGTAGGCAAAAGACCCCTACTGTTATTTCTGTT 1675
Db 1604 AAACAAATTTGACAGTTGAATTTTTCAGACTGCACCTAGGGTACTGTTATTTGCTGC 1663
Qy 1676 TGCTGCAACATTTGCTTCCGCAATATTCAGATCGCGTATTTTCAATGGCCAAATATGG 1735
Db 1664 GGCAGCCACCTTACTGATCTGAACTTCAATGATGCTCGCATAGTGGGCAAAATGG 1723
Qy 1736 CATATTAACATACAGTATGATGATTTTGTGATATCGGTGGTACATCGATGAATTTGAC 1795
Db 1724 TGTCTCAGCAACCGTGTGATGATTTTCTATGATGGTGGAGGATCTGAAAGAGAAATGGA 1783
Qy 1796 CAACCTGATTCATGTTGTTGAAAATGGAATGATGATGTCACAAAGATTTGCTTCCAGA 1855
Db 1784 TAACCTTATAGAAATTTGGTGAAGTGGATCTGATGGGAAGTGGGTACTGTTCCAA 1843
Qy 1856 GCATGTTCCGATTTATTTTATGCAATTAAGAGATGCAATCTGTTGGATTTGGAGATGAAGC 1915
Db 1844 GGAGTTTGAATTTGATTTCTTCTGCACTGCAAGCAGCAGATTTGTGAATAGGAAGAGAGC 1903
Qy 1916 TTTTAAATGCGACGCGGATGTAATAGCCATGTTATTCAACTGTTGGTGAACATAAT 1975
Db 1904 TTTAGTATGGCAGGACGCAATGTTATGAGGAATGTTATCGATGTTGGTGGCTCTGCT 1963

Qy 1976 GAATAGTATGTTGAGAGAGCTATATGGAACAAGAGATGCTTATGTGCCAACATTTAAATGA 2035
Db 1964 GAAGGTGATGAGAAAGAGAGCTGAATGTGTCGACAAATAAGGTAGTGCCTATCAATGGGTGA 2023
Qy 2036 ATATATGAAAACGCTTACGTGCTCATTTGCAATAGCCCGATTTGTCAGCCCGCTATTATA 2095
Db 2024 ATATATGAAACAGCCCATGTATCATTCGCTTGGGACCTATAATCTTCCAAATGCTCTT 2083
Qy 2096 CTTTGTGGGGCCCCAAATATTCAGAGGAGATTTGTTGAAAAGCTCTGAATATCATTAATCTATT 2155
Db 2084 CTTTGTGGACCTAAATCTCTCAGAGGAAATGATTGGAAGCTCTGAATACCAGACTTATA 2143
Qy 2156 TAAGCTATGACGACGCGAGGCTGCACTTCTTAAACATATCCATAGCTTCAAGAGGAAT 2215
Db 2144 TAAGCTATGAGCACTGCTGCTGCTTAAAGAAATGATATTCGATCTTACGATAGAGAATG 2203
Qy 2216 TAAGGAAGGCAATTAATAACGCGGTAGCATTCATTTGAGTAACGAGAGAAAGTGGGAAGT 2275
Db 2204 CAAGAGGAAAGCTGAATATTTCTGCTCTGTGGATGATTGATGGGTGTTAATGTCAAC 2263
Qy 2276 GGAAGAACAGGTGTGAGGAGATGATGATGATTAATAAACAAGAGAGAAAGAAATTAAT 2335
Db 2264 CAAGAGGAGGCCATTTGAAGCAATTAAGGGGATTTTTCAGAGGGCGATAGAGAGCTGCT 2323
Qy 2336 GAAATTAATTTTGAAGAAATGTTAGCATTTGTTCTTACAGCTTGTAAAGATGCAATTTG 2395
Db 2324 GGGGTAGTTTTCAGAGGAGAACACTACAA---TTCCAAGAGCTTGTAAAGATTTGTTCTG 2380
Qy 2396 GAAATGTCCTACGCTGTTGAATTTTTCAGCAAAACGATCAGCGGTTTACTGGAACAC 2455
Db 2381 GAAATGATGTCATGTTGAAATCTATTTTACATGGAAGATGATGGGTACACTTCAATAG 2440
Qy 2456 GATTCTTGATCTGTAAGGACATCAATTTACAACCCGTTGTTGCTGTTG 2502
Db 2441 GTTCATGAACACTGTAAGAGCAATGTTTGAACAAACCCATGATCTGG 2487

RESULT 6

ADM98961

ID ADM98961 standard; DNA; 2658 BP.

XX ADM98961;

XX 01-JUL-2004 (first entry)

XX Diterpene synthase DNA #21.

XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;

XX diterpene precursor; diterpene synthase; defence toxin;

XX volatile defensive signal; pollinator attractant; photoprotectant.

XX Cucurbita maxima.

XX US2004072323-A1.

XX 15-APR-2004.

XX 07-JAN-2002; 2002US-00041018.

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX New unicellular organisms comprising exogenous nucleic acids encoding a

XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for

XX producing diterpenes and diterpene precursors.

XX

| | | | |
|----|--|---|------|
| PS | Claim 3; SEQ ID NO 381; 38pp; English. | | |
| XX | The invention relates to a unicellular organism for producing a diterpene or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polyaromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents DNA encoding a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html. | | |
| XX | Sequence 2658 BP; 757 A; 486 C; 626 G; 789 T; 0 U; 0 Other; | | |
| QY | Query Match | 28.0%; Score 782.2; DB 12; Length 2658; | |
| QY | Best Local Similarity | 61.3%; Pred. No. 2.7e-174; | |
| QY | Matches 1390; Conservative | 0; Mismatches 823; Indels 54; Gaps 6; | |
| QY | 275 | AACTAATCCACCTAATCTGATCATTTGATCAACCAAGAACGATCCAAACAGTTTAA | 334 |
| DB | 236 | AACTAAACCTGGAGCTTTGCACTTTGAAGAAACAAAGAAAGAAATTAAGAAATTTGTCGA | 295 |
| QY | 335 | AAATGTAGAAATTTCTGTTCTTCTATATGACACAGATGGTAGCCATGCTCCCTCTCC | 394 |
| DB | 296 | CHAGGTTGAACCTTTGATTTCTGCATATGATATCTGATGGTGGCAATGGTTCCTCTCC | 355 |
| QY | 395 | AAACTCACCACAAATCGCTTTGTTTCCCTGAGTGTCTCAATTTGGTTAATTAATCAGCT | 454 |
| DB | 356 | AAACTCTCAACCAACCTCTTTTCCCGAGTGTATAAATCTGGGTATTAGATAGTCAACA | 415 |
| QY | 455 | TAATGATGGTTATGGGTCTTGTGTTAATCTACATCTAATCATATCAATCACCCTTGCTTAA | 514 |
| DB | 416 | TGCTGATGGTCTATGGGGCTACTCCACAC-----GATCAGTTGCTGATGAA | 463 |
| QY | 515 | AGATTCTCTATCTCAACATTAGCATGTATTGTTGCATTAAGAAAGATGGAATCTTGGGA | 574 |
| DB | 464 | GGCAATCTCTTATCTACATTAGCATGTGTTCTTACTCTTAAGCGTGGATATTGGGCA | 523 |
| QY | 575 | AGATCAATAATAATAAAGGCTTAAGTTTATTAGTCAAAATCTTGTCTGCTACTGAAA | 634 |
| DB | 524 | TGATCATATAGCAAGCCCTTGATTTTATCAAGTCTAATATAGTCTCAGTACTGATGA | 583 |
| QY | 635 | AGTCAACCATCTCCCATTTGGTTTGGACATCATTTTCTGTTGCTTCCAGTATCGAA | 694 |
| DB | 584 | GAACCAACGTTCTCCGGTGGGATTTGACATATTCTCCCTGGCAAGTATGAGTATGCTAA | 643 |
| QY | 695 | AAACTTGGACATAAACCCTCTTTCAAAACAAACAGATTTTATGTTGATGCTACATAAGAG | 754 |
| DB | 644 | AGACTTGAATTTGAATCTACCTTGGCACCAGCAACGTTGATGCCCTTGGTTCGAAGAA | 703 |
| QY | 755 | GGAAATGGACCAAAAA-----AGATGCCATTTCAATAGATGATGATGATGATGATGATG | 808 |
| DB | 704 | AGAGTTGGAGCTGAGAGAGCTGCAGAACCACTCTGAAGGTGGAAGAGCCCTATTAGCGTA | 763 |
| QY | 809 | TATCTCTGAAGGCTCGGTAAATTTATGATGGAATATGATGGAAGAAATATCAGATGAA | 868 |
| DB | 764 | TGTTTCAGAGGAATTTGAAGTTTACAGGACTGGGATATGGTCTATGCAATATCAAGGAA | 823 |
| QY | 869 | AAATGGTTCTGTTTTCATCACTCACCATCAGCAACAGCTGCTGCTTTCATTAATCATCAAAA | 928 |
| DB | 824 | GAATGGATCACTGTTTAAATTTCCATCAGCAGCGGCTTTTATGTCATAGAAATGA | 883 |
| QY | 929 | TCCTGGTTGCTTAATTAATTTAAATTTCACTTTTGGACAAAGTTTGGTAATGAGTCCCAAC | 988 |
| DB | 884 | TGATGGCTGTTTGTGATTAATCTTCGCTCACTCTTACAAAAGTTTGAAGCTGAGTCCCAAC | 943 |
| QY | 989 | AGTTTATCTCATGATTTTATTTATCCGACTTTCTATGTTGATGACAAATTTGAAAGATTAGG | 1048 |
| DB | 944 | AATATATCTCTGATATATATATGCTGATACATGTTGATAGCTTTCAAAATTCGG | 1003 |
| QY | 1049 | AATTTACACACCACTTTACAGATGGAATTAATAATTTTATAGATGAAACATACAGATGTTG | 1108 |
| DB | 1004 | AATTTCTCGGCATTTTCAAGAGAGAGATAGAGCGTATTAGATGAACTTTACAGGTGTTG | 1063 |
| QY | 1109 | GGTGAACAGATGAGCAAAATTTTATGATGTTTCTTACATGTTTCTTACATGTTTCTGCTT | 1168 |
| DB | 1064 | GATCAAGAGAGAGGAAATATATTTCTTAGATGCTTCACTTGTGCAATGGCTTTTCGAT | 1123 |
| QY | 1169 | ATTAAGGATCAATGGGTATGAAGTTTCCCAATGATCCATTTGGCTGAAATTTACTAA----- | 1222 |
| DB | 1124 | GTATGCTGTTGAAGGATATGATTTTCTTACAGCAGTTGACTCAATTTTTCAGAAATAT | 1183 |
| QY | 1223 | -----TGAATTAGCTTTGAAAGAGATATGAGCTCTTGAACATATCA | 1267 |
| DB | 1184 | CTTTCCCAATTTGCTTTGGAGGATATTTTAAAGACTTCGGTCTGCTGCTGATATATAA | 1243 |
| QY | 1268 | TCGGTCAATATATATATACCAAGAGATTTTATCTTCT---GGAAACAAATCTTGAAGTC | 1324 |
| DB | 1244 | GGCTCTCAGATATATACACGCCCGATGAAATCTGTTCTGGAATATATATACTTTGGAC | 1303 |
| QY | 1325 | AGCTGATTTTCTCAAGAGATATATFCCAATGATTTCAAAACAGCTTTTCTAAA----- | 1376 |
| DB | 1304 | TAGTCTGTTTCTGAGGATATGATTTCTAGTATTCAGTTTGTCTGTAGAACCGATAG | 1363 |
| QY | 1377 | TTAATTTCAAAAGAGTGGAAAAATGCTCTTAAAGTCTTCAATCAATACCGTTTGAACG | 1435 |
| DB | 1364 | TGTTGTTAAACAGAGGCTGTTTAAATGCTCTTGAAGTCTTCCCTATATAATGCAACTGA | 1423 |
| QY | 1436 | CATTAACCTAGACGAAATATACAGCTTTTACAAATGATGACAAATACAAAGATTTCTGAAA | 1495 |
| DB | 1424 | CCTAATATAGTAAGAGGCAATGGAAGTTTACGTGGAGACATTTGTCAGGATTTTCAAAATC | 1483 |
| QY | 1496 | TACATATCACTCATCAATATATTAGTAACACATGATTAACCTTAAGTTTGGCTGTTGAAATTT | 1555 |
| DB | 1484 | GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGGAATTTGCTGTAGAGATTT | 1543 |
| QY | 1556 | CTACACCTGCCAATCTATTTATCGTGAAGAAATTAAGGTTTGAAGGTTGAGGTTGGTGA | 1615 |
| DB | 1544 | CAATACCTGCAACGCAATTCATCTTAAAGAACTTGGAGAGCTTCAAGATGGGTGGTTGA | 1603 |
| QY | 1616 | GAATAGTTGGACCACTCAAGTTTCTGCTAGGCAAAAGACCGCTACTGTTTATTTCTCTGT | 1675 |
| DB | 1604 | AAACAAATTTGAGCAGATTTGAAATTTTTCAGCTGACCTAGGTAATCTGCTATTTTCTGTC | 1663 |
| QY | 1676 | TGCTGCAACACTTTCTGCTCTCCGAATTTATCAGATGCGCGTATTTTCATGCGCCAAAATGG | 1735 |
| DB | 1664 | GGCAGCGACCTTACTGATCTTCACTTCAATGATGCTGCAATAGCATGGGCAACAAATGG | 1723 |
| QY | 1736 | CATATTAACATCAGTATGATGACTTTTGTGATATCGGTGCTACATCCATGCAATTTGAC | 1795 |
| DB | 1724 | TGTGCTCAGACCGTGGTTGATGATTTTCTATGATGTTGGAGGATCTGAAGAGAAATTTGA | 1783 |
| QY | 1796 | CAACCTGATTTCAATGTTTGAATAATGGAATGTAGATGTCCGAAAGGATTTGTTTCTCAGA | 1855 |
| DB | 1784 | TAACTTTATAGAAATTTGGTGAAGAGTGGGATCTGATGGGAAAGTGGTTTACTGTTCCAA | 1843 |
| QY | 1856 | GCAATGTTGGAATTTTATTTTATGCAATTAAGATGCAATCTGTTGATTTGAGATGAGC | 1915 |
| DB | 1844 | GGAGCTTGAGATTTGATTTTCTTGCACCTGCAACAGATTTTGTGAATAGGAAGAGAGC | 1903 |
| QY | 1916 | TTTTAATGCAAGCGCGGATGTAACTAGCCATGTTTATTTCAAACTTTGGTTGGAACTAAT | 1975 |

```

1904 TTTAGTATGCAAGGAGCGCAGTGTATGAGGAATGTTATCGATGGTTGGTCTGCT 1963
1976 GAATAGTATGTTGAGAGAGCTATATGGAACAAGAGATGCTTATGTCACCAATTAATGA 2035
1964 GAAGGTCAVGAGAAAGAGCTGAATGCTCGACAATAAAGTAGTGCCTCAATGGGTGA 2023
2036 ATATATGGAAGAACGCTTACGTGTCAATTCATTTGATAGGCCCGCATTTGCAAGCGGCTATTTA 2095
2024 ATATATGGAACAAGCCCATGATATCAATTCGGTTGGGACCTATATATCTTCCATGCTCTT 2083
2096 CTTTGGGGGCCCCAATTTATCAGAGAGATGTTGAAAGCTCTGAATATCAATATCTATT 2155
2084 CTTTGTGGACCTAAACTCTCAGAGGAAATGATTGGAAGCTGTGAATACCCAGAAGTTATA 2143
2156 TAAAGCTAATGAGCAGCAGCGTGCATCTCTTAAACGATATCCATAGCTTCAAGAGGGAATT 2215
2144 TAAAGCTGATGAGCAGCTGCTGGTGGCTTAAAGATGATATTCGATCTTACGATAGAGAATG 2203
2216 TAAAGGAGCAAAATAAAGCGGTAGCATTTGCAATTTGAGTAACGGAGAAAGTGGGAAAGT 2275
2204 CAAAGAGGGAAGCTCAATATTTCTGTCTCTGTGGATGATTGATGGGGTGTAAATGTCAC 2263
2276 GGAAGAGAGGTTGTGAGGAGATGATGATGATGATTAATAACAAGAGGAAAGAAATTAAT 2335
2264 CAAAGAGGAGGCCAATGGAAGCAATTAAGGGGATTTTGAGAGGGCGGATAAGAGAGCTGCT 2323
2336 GAAATTAATTTTGAAGAAATGGTAGCATTTGCTCTGATGATGATGATGGGGTGTAAATGTCAC 2380
2324 GGGGTAGTTTTCAGAGGAAACACTACAA---TTCCAGAGCTGTGAAGGATTTGTTCTG 2395
2396 GAAATGTCACGTTGTAATTTTTTTTACGAAACGATGACGGGTTTACTGGAACAC 2455
2381 GAAATGATGTCATTTGTAATCTATTTTACATGGAAGATGATGGTACACTTCAAATAG 2440
2456 GATTTCTTATGATGTAAGGACATCATTTACAAACCGGTTGGTGTG 2502
2441 GTTGAAGAACACTGTAAAGCCATGTTTGAACAACCCATGGATCTGG 2487

```

RESULT 7

```

AAT60339
ID AAT60339 standard; cDNA; 2658 BP.
XX
AC AAT60339;
XX
XX
XX 17-OCT-2003 (revised)
DT 27-JUN-1997 (first entry)
XX
XX
DE Pumpkin ent-kaurene B synthase cDNA.
XX
XX Ent-kaurene B synthase; plant growth; seed growth; transgenic plant;
KW antisense; pumpkin; ss.
XX
XX Cucurbita maxima; L cv. Riesenmelone gelb vernetzt.
XX
XX
FH Key Location/Qualifiers
FT CDS 138..2507
FT FT /*tag= a
FT FT polyA_signal 2611..2616
FT FT /*tag= b
XX
XX EP768381-A2.
XX
XX 16-APR-1997.
XX
XX 07-OCT-1996; 96EP-00116005.
XX
XX 09-OCT-1995; 95JP-00261147.
XX
XX (RIKA ) INST PHYSICAL & CHEM RES.
XX
XX Kamiya Y, Yamaguchi S;

```

```

XX
DR WPI; 1997-214817/20.
XX P-PSDB; AAM14055.
XX
XX Pumpkin ent-kaurene B synthase polypeptide - used for the regulation of
XX plant and seed growth.
XX
XX Claim 7; Page 13-19; 24pp; English.
XX
XX A cDNA clone (AAT60339) codes for pumpkin ent-kaurene B synthase
XX (AAM14055), an enzyme that catalyses the conversion of copalyl
XX pyrophosphate to ent-kaurene and which has important functions in the
XX growth of plants and seeds. It is obtd. as a cDNA insert in plasmid pK320
XX following amplification of pumpkin immature cotyledon cDNA using primers
XX based on isolated peptides (see also AAM14056-60) of the purified enzyme,
XX and screening of an immature cotyledon library using the PCR product as
XX probe. The isolated cDNA can be used to produce recombinant kaurene
XX synthase or to produce transgenic plants of increased height. Antisense
XX DNA can be used to produce transgenic plants of reduced height without
XX reducing seed and fruit prodn. (Updated on 17-OCT-2003 to standardise OS
XX field)

```

Query Match 28.0%; Score 780.6; DB 2; Length 2658;

Best Local Similarity 61.3%; Pred. No. 6.4e-174;

Matches 1389; Conservative 0; Mismatches 824; Indels 54; Gaps 6;

```

QY 275 AACTAATCCCCTAATCTGATCATTCATGATACACCAAGAAACGATCCCAAAAACAGTTTAA 334
DB 236 AACTAATAACTGGAGCTTTGCACTTTGAAGAAACAAGAAAGAAATTAAGAAATTTTGA 295
QY 335 AATGTAGAAATTTCTGTTTCTTATATGACACAGCATGGGTAGCCATGGTCCCTTCTCC 394
DB 236 CAAAGTTGAACTTTCAGTTTTCGCATATGATACATGATGATGATGATGATGATGATGAT 355
QY 395 AACTCACCACCAATCGCTTGTTCCTGAGTGTCTCAATTTGTTAAATTAATCAATCAAGT 454
DB 356 AACTCTCTCAACCAACCTCTTTCCCGGAGTGTATAAATCGGTATTAGATAGTCAACA 415
QY 455 TAATGATGTTTCATGGGCTTGTGTTAATCACTCATATCATATCAATCAATCAATCAATCAAT 514
DB 416 TGCTGATGGCTCATGGGCTTACTCCCAAC-----GATCAGTTGCTGANGAA 463
QY 515 AGATTCCTATCTTCAACATTTAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 574
DB 464 GGCCTAATCTCTTATCTATCATTTAGCATGATGATGATGATGATGATGATGATGATGATGAT 523
QY 575 AGATCAAAATAAATAAGGTCTAAGTTTATGATGATGATGATGATGATGATGATGATGATGAT 634
DB 524 TGATCATATGAGCAAGGCCCTTGATTTTATCAAGTCTAATATAGCTTCAAGCTTCAATGA 583
QY 635 AAGTCAACCATCTCCCATTTGGTTTGTGATCATCATATTTCTGTTTGTGATGATGATGATGAT 694
DB 584 GAACCAAGCTTCCGGTGGGATTTGATTTTCCCGGATGATGATGATGATGATGATGATGATGAT 643
QY 695 AAATTGGACATAAACCTCTTTCAAAACAAACAGATTTTATGTTGATGATGATGATGATGATGAT 754
DB 644 AGACTTGAATTTGAATCTACCTTGGCACCGACGACGATGATGATGATGATGATGATGATGAT 703
QY 755 GGAATTGGAGCAAAA-----AGATGCCATTTCAATGAGATGATGATGATGATGATGATGAT 808
DB 704 AGAGTTGGAGCTGAGAAAGCTGCAAGCAACTCTGAAGGTGGAAGGCTATTTAGCGTA 763
QY 809 TATCTCTGAAGGACTCGGTAATTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 869
DB 764 TGTTTCGAAGGATTTGGAAGTTTACAGGACTGGGATGATGATGATGATGATGATGATGATGAT 823
QY 869 AAATGGTTCGTTTCAACTCACCATCAGCAACAGCTGCTGCTTCTTCAATTAATCATCAAAA 928
DB 824 GAATGGATCACTGTTTAAATTTCTCAATCCACTAGCGGAGCGGCTTTTATGATGATGATGAT 883
QY 929 TCCTGGTGTCTTAATTTAATTTAAATTTCACTTTTGGCAAGAGTTTGGTAATGATGATGATGAT 988

```

| | | | | | | |
|----|------|---|------|----------|---|------|
| Db | 884 | TCATGGCTGTTTGAATTAATCTTCGCTCACTCTTACAAAGTTTGAAGCTCAGTCCAC | 943 | 1964 | GAAGTGATGAGAAAGGAAGCTGAATGCTGACAAATAAGGTAGTCCATCAATGGGTGA | 2023 |
| Qy | 989 | AGTTTATCTCATGATTTTATTTATCCGACTTTTCTATGGTTGACAAATTGAAGATTAGG | 1048 | 2036 | ATATATGAAACAGCTTACGTTCTATTTGCAATTTAGCCCGGATTTGCAAGCCGGCTATTTA | 2095 |
| Db | 944 | AAATATATCTCTTGATATATATCTGCTGATATACATATGGTTGATAGCTTCAAAAATCGG | 1003 | 2024 | ATATATGAAACAGCCCAATGATCAATTCGGTGGGACCTATAATCTTCCAATGCTCTT | 2083 |
| Qy | 1049 | AAATTCACACATTTTCAGAGTGGAAATTTAAAATGTTTGTAGATGAAACATACAGATGTTG | 1108 | 2096 | CTTTTGGGGCCCAAAATTATCAGAGGAGATTTGTTGAAAGCTCTGAATATCATATCTATT | 2155 |
| Db | 1004 | AAATTCCTCGGCAATTTCAAGAGGAGATTAGAAGCGTATTAGATGAACCTTACAGCTGTTG | 1063 | 2084 | CTTTTGGACCTAAACTCTCAGAGGAAATGATTTGGAAGCTGTGAATACCAGAGTTATA | 2143 |
| Qy | 1109 | GGTGAACGAGATGACCAATATTCATGATGTTGTAACATGTTTGTAGCTTTTCGGTT | 1168 | 2156 | TTAGCTAATGACACGCGGGTCTGACTTCTTAAACGATATCCATAGCTTCAAGAGGGAAT | 2215 |
| Db | 1064 | GATGCAAGGAGAGGAAAATATATTTCTTAGATGCTTCACTGTCGATGCGCTTCGAT | 1123 | 2144 | TTAGCTGATGACACTGCTGGTGGCTTAAAGAAATGATATTCGATCTTACGATGAGAATG | 2203 |
| Qy | 1169 | ATTAAGGATCAATGGGTGATGAAGTTTCCCGAGATCCATTTGGCTGAAATTAATA | 1222 | 2216 | TTAGGAAGGCAAAATTAACGCGGTAGCAATTCATTTGAGTAACGGHAGAAAGTGGGAAAGT | 2275 |
| Db | 1124 | GTTACGTGTTGAAGGATATGATGTTTCTTCAGACCAAGTTGACTCAATTTTCAGAGATAT | 1183 | 2204 | CAAGAGGGAAGCTGAATTTCTCTCTCTGTTGGATGATTTGATGGCGGTGGTAATGTCAC | 2263 |
| Qy | 1223 | -----TGAATTAGCTTTGAAAGACGATATGCGCTCTTCTGAAACATATCA | 1267 | 2276 | GGAGAAGAGGTTGTTGGAGGAGATGATGATGATGATTAATAAACAAGAGAAAGATTAAT | 2335 |
| Db | 1184 | CTTTTCCCAATGCTTTGGAGGATATTTAAAGACTTTCGGTGCCTCGCTGGAGTTATATA | 1243 | 2264 | CAAGAGGAGGCCCAATTTGAAGCAATTTAAAGGGGATTTTGAGGGCGCATAGAGAGCTGCT | 2323 |
| Qy | 1268 | TGCGTCACATATATATACCAAGAGGATTTATCTTCT---GGAAAAACAATCTTGAAGTC | 1324 | 2336 | GAAATTAATTTTGAAGAAATGGTAGCAATTTGTTCTAGAGCTTTGTAAGAGATGCAATTTG | 2395 |
| Db | 1244 | GGCTCTCAGATATACCGCACCCGAGATGATCTGTTCTGAAATATATAAATCTTTGGAC | 1303 | 2324 | GGGGTTAGTTTTCAGGAGAACACTACAA---TTCCAAGAGCTTTGTAAGATTTGTTCTG | 2380 |
| Qy | 1325 | AGCTGATTTCTCAAGAGATATATCCACTGATCAAAAGAGCTTTCTTAA----- | 1376 | 2396 | GACATGTCACGCTGTTCAATTTTTCGCAACGATGACGGTTCATCTGCGAAACAC | 2455 |
| Db | 1304 | TAGTCGTTTCTGAAGCATGATATCTAGTGATTCAGTTTGGTCTGATAGAACCGATAG | 1363 | 2381 | GAAATTTGATGCTCAATTTGTAATCTATTTTACATGGAAGATGATGGGTACACTTCAAAATAG | 2440 |
| Qy | 1377 | TTAAATTCACAAAGAGTGAAATGCTCTTAAGTTCCCTATCAATACCGTTTGAAGC | 1435 | 2456 | GATTTCTGTACTGTGAAGGACATTTTACACCCGTTGGTGGCTTG | 2502 |
| Db | 1364 | TGTTGTTAAACAAGAGGCTTTAATCCTCTGAGTTCCCTATATAGTCAACTCTAGAACG | 1423 | 2441 | GTTGATGAACACTGTAAAGGCCATGTTTGAACCAACCCATGGATCTGG | 2487 |
| Qy | 1436 | CATAACACTAGACGAATATACAGCTTTTACAATGTAGACAATACAGAATTTCTGAAAC | 1495 | RESULT 8 | | |
| Db | 1424 | CCTAATAAGTAAGAGGCAATGGAAGTTACAGTGGAGACATTTGTGAGGATTTCAAAATC | 1483 | ADM98957 | | |
| Qy | 1496 | TACATATCACTCATCAATATTAAGTAACTAGTACTTAAAGTTGGCTGTTGAGATTT | 1555 | ID | ADM98957 standard; DNA; 2506 BP. | |
| Db | 1484 | GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGGAACTTCTGCTAGAGGATTT | 1543 | XX | ADM98957; | |
| Qy | 1556 | CTACACTGCAATCTATTTATCTGTAAGATTTAAAGGCTTTGAAAGTGGGTGGTGA | 1615 | AC | 01-JUL-2004 (first entry) | |
| Db | 1544 | CAATACCTGCAACGCAATCATCTTAAAGAACTGGAAGAGCTTCAAAGATGGGTGGTGA | 1603 | DT | Diterpene synthase DNA #17. | |
| Qy | 1616 | GAATAAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTGTTTATTTCTCTGT | 1675 | DE | Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene; | |
| Db | 1604 | AAACAAATTTGACAGATGAAATTTTTCAGACTGCACCTAGGGTACTGCTATTTTGTCTGC | 1663 | KW | diterpene precursor; diterpene synthase; defence toxin; | |
| Qy | 1676 | TGCTGCAACACTTTTCGTTCCGCAATATACAGATGCGGATTTTCATGCGGCAAAATGG | 1735 | KW | volatile defensive signal; pollinator attractant; photoprotectant. | |
| Db | 1664 | GGCAGCAGCCCTTACTGATCTGAACTTCATGATGCTCGCATAGCATGGGCACAAAATGG | 1723 | OS | Arabidopsis thaliana. | |
| Qy | 1736 | CATATTAACACTAGTATGATGATTTTGTATCGGTGATCAATCGATGATTAATGAC | 1795 | PN | US2004072323-A1. | |
| Db | 1724 | TGTTCTCAGCAGCCGTTGATGATTTCTATGATGGTGGAGATCTGAAGAGGAATTTGA | 1783 | PD | 15-APR-2004. | |
| Qy | 1796 | CAACCTGATTCATGTTGAAAAATGGAATGTAGATGTCACAAAGGATTTCTGTTTCA | 1855 | PF | 07-JAN-2002; 2002US-00041018. | |
| Db | 1784 | TAACTTATAGAAATTTGTAAGATGGATCTGATGGGGAAGTGGGTACTGTTCCAA | 1843 | PR | 05-JAN-2001; 2001US-0259880P. | |
| Qy | 1856 | GCATGTTGCGATTTTATTTTATGCAATTAAGATGCAATCTGTTGGATTTGAGATGAGC | 1915 | XX | (MATS/) MATSUDA S P T. | |
| Db | 1844 | GGAGTTGAGATTTGATTTCTTGACACTGCACAGCAGCTTTTGTGAAATAGGAAGAGC | 1903 | PA | (HART/) HART E A. | |
| Qy | 1916 | TTTTAAATGGCAGCGCGATGTAACCTAGCCATGTTATTCAACTGTTGGTGGACTAAT | 1975 | PI | Matsuda SPT, Hart EA; | |
| Db | 1904 | TTTAGTATGGCAAGCAGCGATTTATGAGGAATGTTATCGATGTTGGTGGTCTGCT | 1963 | XX | WPI; 2004-373921/35. | |
| Qy | 1976 | GAATAGTATGTTGACAGAGCTATATGGAAGAGATGCTTATGTCACCAATTAATGA | 2035 | XX | New unicellular organisms comprising exogenous nucleic acids encoding a | |
| | | | | PT | geranylgeranyl pyrophosphate and a diterpene synthase, useful for | |
| | | | | PT | producing diterpenes and diterpene precursors. | |
| | | | | PS | Claim 3; SEQ ID NO 377; 38pp; English. | |
| | | | | XX | The invention relates to a unicellular organism for producing a diterpene | |

or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polycyclic aromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents DNA encoding a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 2506 BP; 760 A; 454 C; 586 G; 706 T; 0 U; 0 Other;

Query Match 26.2%; Score 732.4; DB 12; Length 2506;

Best local similarity 59.5%; Pred.NO. 1.5e-162;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

Sequence 2506 BP; 760 A; 454 C; 586 G; 706 T; 0 U; 0 Other;

Query Match 26.2%; Score 732.4; DB 12; Length 2506;

Best local similarity 59.5%; Pred.NO. 1.5e-162;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

| | | | |
|----|-----|---|-----|
| QY | 270 | GGACAAACTTAATCCCACTAATCTGATCATTTGATACAACCAAGAACGGATCCAAAACAG | 329 |
| Db | 143 | GTA CAGACA GAGCTAA CAATGTGAGCTT TGAGCAAACA AAGGAGA AGATTAGGAAGTG | 202 |
| QY | 330 | TTTTAAAAATGTAGAAATTCGTGTTCTTCATATGACACAGCATGGGTAGGCCATGGTCCCT | 389 |
| Db | 203 | T TGGAGAAAGTGGAGCTTTCGTGTTCGGCCCTACGATACTAGTTCGGGTAGCAATGGTTC | 262 |
| QY | 390 | TC TC CCAACTCACC CCAATGCCTTGTGTTCCCTCGAGTGTCTCAATGGTGTAAATTAATAAT | 449 |
| Db | 263 | TC ACCGAGCTCC CAAAATGCTCCACTTTTCCCACAGTGTGMAATGGTTATGGATTAAT | 322 |
| QY | 450 | CAGCTTAATCATGGTTCATGGGCTCTGTGTTAATCACACTCATATCATATCACCCGGTGG | 509 |
| Db | 323 | CAACATGAAGATGATCTTGGGGACTTGATAACC-----ATGCCATCAATCTCTTT | 373 |
| QY | 510 | CTTAAAGATTCTCTATCTTCAACATTAGCATGTATGTGCTATTA AAAAGATGAAATCTT | 569 |
| Db | 374 | AAGAAGGATGTGTTATCATCTACACTCGCTAGTATCCTCGGTTTAAAGAGTGGGGAATT | 433 |
| QY | 570 | GGGGAGATCAAAATAATAAAGTCTAAGTTTAAATGAGTCAAACTTGCTTCAGCTACT | 629 |
| Db | 434 | GGTGAAGAACAATAAACAAGGGTCTCCAGTTTATTGAGCTGAATCTCGCATTAGTCACT | 493 |
| QY | 630 | GA AAAAAGTCAA CCATCTCCCATTTGGTTTGGACATCATATTTCTCGGTTGCTTGA | 689 |
| Db | 494 | GATGA AACCATACAGAAACCAA CAGGTTTGTATTTATTTCTCGGGATGATTAATAT | 553 |
| QY | 690 | GGCAAAAAC TTGGACATAAACCTCTTTCAAAA CAACAGATTTTATGTTGATGCTACAT | 749 |
| Db | 554 | GCTAGAGATTGAATCTGACGATTC CATTCGAGTGGGCTCAGAAGTGGTGGATGACATGAT | 613 |
| QY | 750 | AAGAGGGAATTGGAGCAAAAAG-----ATGCCATTCAAATGAGATGGATGATAC | 800 |
| Db | 614 | AAAAGGATCTGGATCTTAAATGTGATGTGA AAAAGTTTCAAAGGGAAGAGAAGCATAT | 673 |
| QY | 801 | TTGGCTATATCTCTCAAGGACTCGGTAAATTTATATGATTGGAATATGGTGAAGAAATAT | 860 |
| Db | 674 | CTGGCCTATGTTT TAGAGGGGACAAGAAACCTTAAAGATTGGGATTTGATAGTCAAAATAT | 733 |
| QY | 861 | CAGATCAAAAATGGTTCCTGTTTTCACCTACCATCAGCAACAGTGCTGCTTTCAATTAAT | 920 |
| Db | 734 | CAAAGGAAATGGGTCACGTGTTTGATTCTCCAGCCACAACAGCAGCTGCTTTTACTCAG | 793 |

QY 753 AAGCAATCGAGCAAAAAGATGCC-----ATTCAAATGAGATGGATGATCTTG 803
Db |||||
QY 312 CGGAGATGGAATTTGAAAAGCTGGCTGTGGATAGTCTTTTGGAGAAAAGCHATATG 371
Db |||||
QY 804 GCGTATATCTCGAAGCACTCGGTAAATTTATATGATGGAAATANGTGGAAGAAATATCAG 863
Db |||||
QY 372 GCTTTTATCCCAAGAGGATTCGGAATATGCTGCACTGGGATCAAGTTATGAAGTTTCAG 431
Db |||||
QY 864 ATGAAAAATGGTCTCTCTTTTCAACTCACCATCAGCAACAGCTGCTCTCTTTCAATTAATCAT 923
Db |||||
QY 432 AGAAGAAATGGATCATGTTTCAGCACTCTCTTCCACAACACTGCTGTGCATTAATFCCAATA 491
Db |||||
QY 924 CAAAATCTGGTGTCTTAATTTATTAATTTCACTTTTGGACAAGTTTGGTAATGCACTC 983
Db |||||
QY 492 TACAACGACCAAGCCCTTCAATACCTAAATTTGCTTGTCAATGCAATTTGGCAGTGCACTA 551
Db |||||
QY 984 CCAACAGTTTATTCCTCATCATGATTTATTTATCCGACTTTTCTATGTTGACACAATTTGAAGA 1043
Db |||||
QY 552 CCAGCAATGTATCTCTCAAGGGTACATTTGTCAGCTTTCATGTTGGACGCGCTTGAAAAA 611
Db |||||
QY 1044 TTAGGAATTTACACCAATTTCAGAGTGGAAATTTAAAAATGTTTATAGATGAACATACAGA 1103
Db |||||
QY 612 ATGGGAATTTTCAGCGCTTTGTCAGTGAATAGAAAGCATCTCTGGACATGGCATACAAT 671
Db |||||
QY 1104 TGTGGTGGAAACGAGATGAGCAAAATATTTCATGGATGTTTGTAAACATGCTCTTTAGCCTTT 1163
Db |||||
QY 672 TGCTGTTACAGATGATGAGGAATCTATGATGGACATAGCAACATTTGCAATGGCAATTT 731
Db |||||
QY 1164 CGGTTATTAAGGATCAATGGGTATCAAGTTTCCGAGATCCATTTGGCTGAAATTAATAAT 1223
Db |||||
QY 732 CGCCTTTTGGAGATGAATGGTTACGATGTTTCTCAGATGAGCTGCTCACGTTGCTGGA 791
Db |||||
QY 1224 G-----AATTAGCTTTGAAAGACGAATATGCAGCTCTTGA 1259
Db |||||
QY 792 GCTTCCACTTTCCATGATTCATCAGGATTAATTAATGATACAAAATCCCTACTGGAA 851
Db |||||
QY 1260 ACATATCATGGTCCACA---TATATTATACCAAGAGGATTTATCTTCTGGAAAAACAATC 1316
Db |||||
QY 852 TTGTACAAGACCTCAAAAGTCACTTATCAGAAAACGATCTGATCTTAGATCGCATAGT 911
Db |||||
QY 1317 TTGAAGTCAGCTGATTTCTCTCAAGAGATATATCCAC----- 1354
Db |||||
QY 912 TCCTGGTCTGGCACTTATTGAAGGATGAAGTGTCTGTAGTAGGTGCAAAAGACTCGA 971
Db |||||
QY 1355 -----TGATTCAAAACAGGCTTTCTAAATTAATTCACA----- 1386
Db |||||
QY 972 TTTTGGAGAGATGCTGCAACAAAATTTAAATTTCTCATTTTCACCTTGGAGGTTCAAGT 1031
Db |||||
QY 1387 -----AAGAGTGGAAAATGCTCTTAAGTTCCCTTATCAATACCGGT 1427
Db |||||
QY 1032 TATGTCGTGTTTGTGNTTTTCAGATCGAGTATGCTGTTAATTTCCCTTGTATTCACA 1091
Db |||||
QY 1428 TTAGAAGCATAAACAACATAGACGAAATATACAGCTTTTACAATGTAGACAAATACAGAAT 1487
Db |||||
QY 1092 CTGAGCGCTTAGAACACAAGAGAAAACATCAACATTTTGAATGCTTGGGGTCTCTGATG 1151
Db |||||
QY 1488 CTGAAAACATACATATCACTCATCAAAATATTAGTAACACTGATTACCTAAGGTTCGCTGTT 1547
Db |||||
QY 1152 CT---AACAACAATCTCATCTTTTCTGATCAATCAAGAAATTCCTAGCTTTGGCAGTC 1208
Db |||||
QY 1548 GAAATTTTACACTGCGCAATCTATTATTCGTGAAGAAATTAAGGCTTTGAAAGGTGG 1607
Db |||||
QY 1209 GAAGATTTCACTTCTCTCAACGTGTTTACCGGATGAATTCGCGCATCTTGATAGTTGG 1268
Db |||||
QY 1608 GTGCTAGAGATATTTGGACCACTCAAGTTTCTAGGCAAAAGACCGCTACTGTTAT 1667
Db |||||
QY 1269 GTGAAGAGAACAAAGCTGGACCACTCAATTTCTCGGCAGAACTGACATATGCTAT 1328
Db |||||
QY 1668 TTCTCTGTGTGCAACACTTTGCTTCCGAAATATACAGATGCGGTATTTTCATGGCC 1727
Db |||||
QY 1329 CTGCTCTGCTGTACGGTAATTTCTTCTGAATTTGCTGACGCTCGCATTTTCATGGCC 1388
Db |||||

QY 1728 AAAAAATGSCATATTAACTACAGTAGTTGATGACTTTTGTGATATCGGTGTTACAATCGAT 1787
Db |||||
QY 1389 AAAAAATGGTGTCTCTCAAACTGTGGTTGATGACTTCTTCGATGTTGGTGGATCAAAAGAA 1448
Db |||||
QY 1788 GAATTCACCAACCTGANTCAAATGCTGTTGAAAAATGGAATGTAGATGTCGACAGGATTTGT 1847
Db |||||
QY 1449 GAATTAGAAAACCTGATAGACATAGTTGAGAAATGGCATGGGCACCACTGACGTGAGTTT 1508
Db |||||
QY 1848 TGTTACAGACATGTTCCGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGGATTGGA 1907
Db |||||
QY 1509 TATTCGAAACAGCTGAAAAATAGTATTTTCTGCTATTTATCAACAGTGAACCATCTTGA 1568
Db |||||
QY 1908 GATGAAGCTTTAAATGGCAAGCGGCGGATGTAACTAGCCATGTTATTCAAACTTGGTTG 1967
Db |||||
QY 1569 GCAATGGCTTTCGACGACACAGGCGGTGATCTTCAAAACCACTAGTAGAAAATATGGCTG 1628
Db |||||
QY 1968 GAACTAATGAATAGTATGTTGAGAGAGCTATATGGACAAGAGATGCTTATGTCGCAACA 2027
Db |||||
QY 1629 GATTTGTTAAGATCTATGATGTTGCGAGGAGCAATGGCAGAGATGCCAATATGTACCAACA 1688
Db |||||
QY 2028 TTAATGAATATATGGAACCAACGCTTACGTTGTCATTTGCAATTTAGGCCCGATTTGTCAGCG 2087
Db |||||
QY 1689 GTTGAAGAATACATGACAAATGCTGTTGCTCTCATTTGCACTGGGCCCAATTTGCTCCCA 1748
Db |||||
QY 2088 GCTATTTTCTTTGTCGGGCCCAAAATATCAGAGGAGATTTGTTGAAAGCTCTGAAATATCAT 2147
Db |||||
QY 1749 GCAATTTGATTTTGTAGGGCAAGAGCTATTAGAGCATGCTGTCAAAAGATGAAGATACGAT 1808
Db |||||
QY 2148 AATCTATTTAAGCTAATGAGCACGCGGTCGACTTCTTAAACGATATCCATAGCTTCAAG 2207
Db |||||
QY 1809 AAATTTATTTAGCTAGTAGGACACTTTCGGGGAGCTTCTCAATGACTTACCAAAAGTTTAGAG 1868
Db |||||
QY 2208 AGGGAATTTAGGAAGGCAAAATTAACCGGCTAGCAATTTGCAATTTGAGTAACGGAAGAACT 2267
Db |||||
QY 1869 AGGGAAGGCAACCGGGGAAAGCTGAATAGTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1928
Db |||||
QY 2268 GGGAAAGTGGGAAGAGAGAGTTTGTGAGGAGATGATGATGATGATTTAAAAACAAGAGAAA 2327
Db |||||
QY 1929 TCTATGTCATAGAACCGCTTAAAGGCAATGCAAGAGTCCATAGACGTGTCTAGGAGA 1988
Db |||||
QY 2328 GAATTAATGAATTAATTTTGAAGAAATGTTAGTACTGTTTCTTCTTCTTCTTCTTCTTCTTCT 2387
Db |||||
QY 1989 GACTTGTCTAAGATTTGGTTCT---CAGGAAAGAAAGTGTCTGTTCTTCTTCTTCTTCTTCT 2045
Db |||||
QY 2388 GCATTTGGAAACATGTGTGACGTTTGAATTTTTCACGAAACGATGACGGGTTTA 2445
Db |||||
QY 2046 CTCTTCTGGAAGATGTGTAAGATACTTCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2103
Db |||||

RESULT 10

ADA70130
ID ADA70130 standard; DNA; 2154 BP.

XX ADA70130;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 3453.

XX Plant; bacterial infection; fungal infection; viral infection; rice;
gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
DR Identifying at least one gene involved in plant resistance or response to
XX pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX Claim 6; SEQ ID NO 3453; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interraction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2154 BP; 596 A; 465 C; 556 G; 537 T; 0 U; 0 Other;
Query Match 14.9%; Score 416.8; DB 8; Length 2154;
Best Local Similarity 52.7%; Pred. No. 4.9e-88;
Matches 1109; Conservative 0; Mismatches 937; Indels 60; Gaps 7;
QY 381 ATGGTCCCTCTCCAACTCAACCAATCGCTTGTTCCTCGAGTGTCTCAATTTGGTTA 440
DB 1 ATGGTCCCACTCCGGGGTTCGTCTCAATCCAACTCCAAAGCTTCCCTCATGCGTGGCTGAYA 60
QY 441 ATTAATAATCAGCTTAATGATGGTTTCATGGGGTCTTTGTTAATCACACTCATATATCAAT 500
DB 61 CTAGAGATCAATGGGACGATGATCATGAGTATTGACGGGTCCATCCACAGCCAA- 119
QY 501 CACCCGTTGCTTAAAGATTCCTATCTTCAACATAGCATGTATTTGTTGATTTAAAGA 560
DB 120 -----CAAAGATGTCCTATCATCTACGCTAGCATGTGTCTGGCACTCAACAA 168
QY 561 TGGAAATGTCGGGAGATCAATAAATAAAGGTCTAAGTTTATTTAGTCAAACTTGTCT 620
DB 169 TGGAAATGTCGGGAGATCAATAAATAAAGGTCTAAGTTTATTTAGTCAAACTTGTCT 228
QY 621 TCAGCTACTGAAAAAAGTCAACCACTCCCAATGTTGTTGACATCATATTTCCCTGGTTG 680
DB 229 ATTGGATGATGATCAAGCAGTGTCTCTATAGTTTTCGGCATCACTTTCCAGCTATG 288
QY 681 CTGAGATGATGAAAAAATTGGGACATTAACCTCTCTTCAAAACAAACAGATTTTAGTTG 740
DB 289 CTAACCCCTTGCCAAATGGTCCGGCTTGGAAATGTCCTGTCAGACAAAACGATATTGATAGT 348
QY 741 ATGCTACATAGAGGGAATTCGAGCAAAAAGA-----TGCCATTCAAATGAGATG 791
DB 349 CTTAATCACCTTCGGGAGATGAATAATCAAGAGAGCGCGGGAACCAATCTCTGTTGGA 408
QY 792 GATGGATCTTGGGGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGGT 851
DB 409 AAGGCCATATGGCTATCTGCGAGAGGATTTGGGAACCTTCTGGAGTGGGATGAATC 468
QY 852 AAGAAATATCAGATGAAAAAGGTCTCTTTTCAACTCACTCAGCAAGCAAGCTGCTGCT 911
DB 469 ATGATGTTCCAGAGGAAAGCGGGTCATTGTTCAACTCTCTTCAACTGCTGGCCCA 528
QY 912 TTCAATTAATCAATAAATCTGGTGTCTTAATTAATTTAAATTTCACTTTTGGACAAGTTT 971
DB 529 TTAGCCCAATACCAGCAGATTAAGCTCTCCAACTATTGCAATCTCTAGTCAATAAATTT 588
QY 972 GGTAAATGAGTCCCAACAGTTTATCCTCATGATTTATTTATTCGAGCTTTTATGTTGAC 1031
DB 589 GACGCTGTAGTGCCAAACACTGTATCCACTAAACATATATTGTTCAGCTTTCAATGTTGAT 648

QY 1032 ACAATTGAAAGATTAGGAATTTCAACCAATTTACAGATGGAATTTAAATAATGTTTTAGAT 1091
DB 649 GCGCTTGAACATCGGAATATCTCAGTATTTTGGCCAGTGAATAAAGAGCATCTTGAC 708
QY 1092 GAAACATACAGATTTGGGTGGAACGAGATGAGCAAAATATTCATGATGTTTAAACATG 1151
DB 709 ATGACATACAGTTCTCGTTTAGGAAAGATGAGAAATCATGTTAGACGTAAACAAGTGC 768
QY 1152 GCTTTAGCTTTCGGTTATTAGGATCAATGGGTATGAAGTTTCCACATCATCTGGCT 1211
DB 769 GCAATGGCAATTTCCGCTTTTAAAGAAATGAACGGTTACGATGTTTCTCAGATGAGCTATCT 828
QY 1212 GAAATTAATAAGAAATTAG-----CTTTGAAAGACGAAATAT 1247
DB 829 CATGTTGCTGAGCTTCAGGCTTCGCTGATTCGTTCAAGGGTATCTAAACGATAGAAA 888
QY 1248 GCAGCTCTTGAACATATCATCGCTC---ACATATATTAACAAGAGATTTATCTTCT 1304
DB 889 TCTGTACTGGAAGTATACAAAGACCTCGAAACACAGTATATCAGAAATGATTTGATCTTA 948
QY 1305 GGAACAAATCTTGAAGTCAGCTGATTTCTCAAGAGATAATATCCACTGATTTCAAAC 1364
DB 949 GATAGCATAGGCTCTTGGTCAGGACCCATTTCAAGGAAATGCTGTCTTAATGGGAAA 1008
QY 1365 AGGCTTTCTAAATTAATTCACAAAGAGGTGGAATGCTCTTAAGTTCCCTATCAATACC 1424
DB 1009 GGAATCTC-----AGGCGGTGAAGAGATCGAGTTTGTCTGAAGTATCCCTTCTACAGC 1062
QY 1425 GGTTTAGACGCAATAAACAACACTAGACGAATATACACTTTCAATGTAGACATAAAGA 1484
DB 1063 ACATTGGAACGCTAGTCCACAGGAAACATTTGTCTTTTCGACGCTAAGGCTCCAG 1122
QY 1485 ATCTGGAACACTACATATCACTCATCAAAATATTAGTAAACACTGATTAACCTTAAGTTGGCT 1544
DB 1123 ATGCTGAAACAGAGTGCATGCTCTGTTCAACAGTCA---AGATTTCTAGCTTTGGCT 1179
QY 1545 GTTGAAGATTTTCAACCTGCCAATCTATTATTCGTGAAGAAATTAAGGCTCTTGAAGG 1604
DB 1180 GTGATGATTTCTGCATCTCTCAATCTAATTAACAGAAATGAATGAATTTATCTCGAAAGT 1239
QY 1605 TGGGTGTGAGAAATTAAGTTGGACCACTCAAGTTTGTAGGCAAAAGACGGCTACTGT 1664
DB 1240 TGGGTGAAACACACAGACTCGACAGCTACATTTCCACGGCAGAAAGATAACATATGC 1299
QY 1665 TATTTCTCTGCTGTCGCAACACTTTGCTCTCCGAAATTTACAGATGCGCGTATTTCAATGG 1724
DB 1300 TATCTCTGCTGCTGCCACCACTCCGCTCTGAAATGGGCTACGCTCGACCTCGTGG 1359
QY 1725 GCCAAAAATGGCATATTAACTACAGTATGTTGATGACTTTTTTGTGATATCGGTGGTACATC 1784
DB 1360 GCAAGAACCGCTTGGTTGACGGCTGTTATCGAAGATCTCTTCGATGTTGGTGGATTAGAA 1419
QY 1785 GATGAATGACCAACCTGATTCATGTTGTGAATAAAGTGTAGATGTCGACAAGAT 1844
DB 1420 CAAGAACAGAAACCTCTAGCATTAATGGAGAAGTGGGAAGAGCCTGTGTAAGATGAG 1479
QY 1845 TGTGTTTCAGAGCATGTTCCGATTTTATTTTAGCATTTAAAGATGCAATCTGTGGATT 1904
DB 1480 TACTACTCTGAGATGTAAGATCGTGTTCAGGCTCTTTTATAATAACGGTGAACGAGATT 1539
QY 1905 GGAGATGAAGCTTTTAAATGGCAAGCGCGGATGTAATCTAGCCATGTTATTTCAACTTGG 1964
DB 1540 GCGCAAGGCTTCCGGGTTTACAAGGCCATGATGTTTACCAAGTACCTTAGTAGACGTATGG 1599
QY 1965 TTGAACTAATGATATGATGTTGTAGGAGAGCTATATGGAACAAGATGCTTATGTGCCA 2024
DB 1600 CTAATGTGGGAGGTGATGAAGTTGAGGCAAAATGGCAGAGAGCCAGCACTGCCA 1659
QY 2025 ACATTAAATGAATATATGGAACCGCTTACGTGTCAFTTGCATTTAGGCCCGGATTTGTAAG 2084
DB 1660 ACGTTGAGAAATACATGGAAGTGGCATGTTGTCACTTGGACAGGGCTGTACGGTGATG 1719

QY 2085 CCGGCTATTATTCTTGTGGGGCCCAAAATATATCAGAGAGATTGTTGAAAGCTCTGAATAT 2144
Dd |||||
Dd 1720 TCGGCATCTTCTGATCGGAGAAAAGCTCCGGAGGGTATTCTCGAATTGAGATATC 1779
QY 2145 CATAATCTATTAAAGTAATGACGACGCGAGGTCGACTTCTTAAACGATATCCATAGCTTC 2204
Dd |||||
Dd 1780 GATGAGTTCTTACGGCTGATGGGCACTTGGCGCGCTCTCTGAAATGACATTCGAGGCATT 1839
QY 2205 AAGAGGGAATTTAAGGAAGCAAAATTAAGCGCGGTAGCATTTGATTTGAGTAACGGAGAA 2264
Dd |||||
Dd 1840 GAGAGGAGAGTCGAGCGGCAAAATGACGACGCGCTCTCGTCTCGTTACCGTAGC 1899
QY 2265 AGTGGAAAGTG--GAAGAAGAGGTTGTGGAGAGATGATGATGATTTAAACAAAG 2321
Dd |||||
Dd 1900 GCGGCTCCATGTCCGTAGACGAGGCCAAAACCGAAGTGTAGCGCATCGAGCCTCA 1959
QY 2322 AGGAAGAAATTAATGAAATTAATTTTGAAGAAATGATGATGATTTCTAGAGCTTGT 2381
Dd |||||
Dd 1960 CGGAGGAAGCTGCTGAGCTTGGTCTGTCAGCGAACAGGAAGGCCCTATCCGAGCGCTGC 2019
QY 2382 AAAGATGCAATTTTGGAAACATGTCTACGNGTTGAATTTTACGCAACGATGACGG 2441
Dd |||||
QY 2020 AGCAGCTCTTCTGGAAGATGTGCAAGATCTTCACCTGTTCTACTACCAGACCGGA 2079
Dd |||||
Dd 2442 TTACT 2447
Dd 2080 TTCACT 2085

RESULT 11

AD69971
ID AD69971 standard; DNA; 2223 BP.

AC AD69971;
XX 20-NOV-2003 (first entry)
XX Rice gene, SEQ ID 3294.
DE Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
XX Oryza sativa.
XX
XX WO2003000898-A1.
XX
XX 03-JAN-2003.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.

Identifying at least one gene involved in plant resistance or response to
pathogenic infection for conferring resistance or tolerance to a plant to
bacterial, fungal or viral infection by determining or detecting plant
gene expression.

Claim 6; SEQ ID NO 3294; 899pp; English.

The present invention relates to a method (M1) for identifying genes
involved in plant resistance or response to pathogenic infection. M1
comprises identifying a gene whose expression is significantly altered in
the incompatible interaction of plant gene expression relative to
the expression of the gene in an uninfected plant, in a mutant plant that
does not express a gene associated with response to pathogenic infection,
or in a corresponding incompatible or compatible interaction. (M1) is

CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.

XX

SQ Sequence 2223 BP; 552 A; 528 C; 634 G; 509 T; 0 U; 0 Other;

Query Match 13.6%; Score 379.4; DB 8; Length 2223;
Best Local Similarity 54.5%; Pred. No. 3.4e-79;
Matches 900; Conservative 0; Mismatches 676; Indels 75; Gaps 4;

QY 795 GGATATCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTGGATATGGTGAAG 854
Dd |||||
Dd 574 GAATATACGGCTTATGCTCTCAGAGGATTAGCAATATTCAGACTGGATGAAGTATG 633
QY 855 AAATATCAGATGAAATAATGGTCTCTTTCACTCACCATCAGCAACAGCTGCTGCTTC 914
Dd |||||
Dd 634 AAATTCGAAGAAAGAAATGGCTCACTGTTCAACTCCCTTACACAACACTGCAGCTCATTA 693
QY 915 ATTAATCATCAAAATCCCTGGTGTCTTAATTTAAATTCATTTTGGACAAGTTGGT 974
Dd |||||
Dd 694 GTCCACAATATGATGCCAAAGCTCTCCAGTACTTAGACATGCTCTGACCAAAATTTGA 753
QY 975 AATGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGTTGTCACCA 1034
Dd |||||
Dd 754 AGTGCAGTGCAGCGGCTATCTGCAATATTCAGTCTCAGCTCTACAATGGTATG 813
QY 1035 ATTGAAGATTAAGAAATTTTACACCATTTTCAGAGTGGAAATTAATAATGTTTATGATGA 1094
Dd |||||
Dd 814 CTTGAAAAGATGGGAATATCTAGGCATTTTGTGGTGAATAAAGAGCATACTGGACATG 873
QY 1095 ACATACAGATGTTGGTGGAAACGAGATGAGCAAAATATTCATGATGTTGTACATGCTCT 1154
Dd |||||
Dd 874 ACCTACAGTGTCTGGAACAGAGGATGAGGAAATTTGCTTGACATGCAAAACATGTTGG 933
QY 1155 TTAGCTTTTCGGTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATGGCTGAA 1214
Dd |||||
Dd 934 ATGGCATTTGGTATGTTGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 993
QY 1215 ATTACTAATGAATTTAGCTTTTGAAGACGAATGATGATGATGATGATGATGATGATGAT 1274
Dd |||||
Dd 994 TTTTCT-----GAACCTTCAAGTTTCCCACAATTC 1023
QY 1275 CATATATTATCCAAAGAGATTTATCTTCTGGAACAAATCTTGAAGTCACTGATTTTC 1334
Dd |||||
Dd 1024 CTTCAAGATATCTGAATGATACAAAGATCTTTATTAAGAAATACATAAGGCTTCAAAAGTC 1083
QY 1335 CTCAAAGAGATAATATCCACTGATTCAAACAGGCTTTCTAAATTAATTCACAAAGAGGTG 1394
Dd |||||
Dd 1084 -----AGTATCGCAGAAAGAGGTT 1104
QY 1395 GAAATGCTCTTAAAGTTCCTATCAATACCGGTTTGAACCGAATAAACACTAGACGAAT 1454
Dd |||||
Dd 1105 GAATATGCTCTTGAATTTCCCTTCTATACCATCTTGGATCGTCTAGACCAATAAAGAAAT 1164
QY 1455 ATACAGCTTTCAATGATAGCAATACAGAAATCTTGAAGAACTGAAACATATCACTCATCAAT 1514
Dd |||||
Dd 1165 ATCGAACATTTTGACATTAACAGCAGTCAAGATGCTAGAGAAACAGCGT---ACTTGCATGT 1221
QY 1515 ATTAGTAACTGATTTACTAAGGTTGGCTGTGTAAGATTTCTACACCTGCCAATCTATT 1574
Dd |||||
Dd 1222 CATTCGAATGAAGAAATCATGCGCTTGGGTGTGAGAGATTTTAGTACTCTCAGTTTATT 1281
QY 1575 TATCGTGAAGAAATTAAGGTTTGAAGGTTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1634
Dd |||||
Dd 1282 TTCCAAGAAAGAGCTGCAGCAACTCAACAGCTGGGTGAAAGAGAGCAGGTGGATCAGCTG 1341
QY 1635 AAGTTTGTAGCAAAAGACCGCTACTGTTATTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1694
Dd |||||
Dd 1342 CAATCGCAAGGCAAGATTTGGACTACTTCTATTTCTCTGTGTGTGTGTGTGTGTGTGTGT 1401
QY 1695 CCCGAAATTTATCAGATGCGGCTATTTTCATGGGCCAAAATGCAATTAATCACTACAGTGT 1754
Dd |||||
Dd 1402 CCTGAACGTGTCAGATGTTGGCATTTTGTGGGCCAAAATGCGGTGTGTGTGTGTGTGTGTGT 1461

1755 GATGACTTTTGGATATCGGTGATCAATCGATGATGACCAACCTGATTCATATGCTT 1814
1462 GACGACTTCTTCGACGTTGGAGATCAAAAGAGAACTCGTGGCAATAGTT 1521
1815 GAAATGGAATGTAGATCGCAGAGGATTTGTTTCAGAGCATGTTGCGATTTTATTT 1874
1522 GAGAGTGGGACAAGATGACAAACTGAGTACTCTCTGACAACTAGAGATGTTGTT 1581
1875 TTAGCATTAAGATGCAATCTGTTGGATGAGATGAAAGCTTTTAAATGGCGAGCGCG 1934
1582 TCTGCAATTTATATCTTCAACTAACCACTTGGATCAATGGCCTCTCTAGTACAGCCGT 1641
1935 GATGTAACAGCATGTTTCAAACTGTTGGTGGAACTAATGATAGTATGTTGAGAGAA 1994
1642 GATGTCACCAACACCTTGTAGAAATATGGCAAGAAATGCTGAGGCTATGATGACAGAG 1701
1995 GCTATATGCAACAGAGATGTTTATGTCACCAACATTAATGAATATATGAAACGCTTAC 2054
1702 GTAGAGTGGAGGAGAGCGGTATGTCACACAGCAGAGGAATACATGGAATGCAATT 1761
2055 GTGTCATTTGCAATGAGCGGATGTCACCGCGCTATTTACTTTGTTGGGGCCCAATTA 2114
1762 GTGACATTTGCACTGGGACCGGTTGCTCCAGCATTTGATCTTGTGGACCAAGATC 1821
2115 TCAGAGGAGATTTGAAAGCTCTGAATATCATATTAATTAAGCTAATGACACGCGAG 2174
1822 CCGTATCTGTCATAAGAACCAAGAAATGCGAGGTTGTTCCGGCTTAATGACAAATGT 1881
2175 GGTGCGACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGAAGCAAAATTAAC 2234
1882 GCGCGTCTCTGATGATGTCATCTACGAGAGAGGGCGAGCCAGGCAAGCTGAAC 1941
2235 GCGGTAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2294
1942 AGCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2001
2295 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2354
2002 CAGATTCAGAGACCCATCGAATATGAGAGAGAGATGATGATGATGATGATGATGATGAT 2058
2355 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2414
2059 AGAGGAGGCGCGTTCACAGGCCATGAGGAGGCTGTTCTGGAGCATGTCGAAGTCTGC 2118
2415 AATTTTATGCAAAACGATGACGGTTTA 2445
2119 CACTTCTTCTACTCCGCGCGGACGGGTTCA 2149

RESULT 12

AD145176
ID AD145176 standard; cDNA; 2223 BP.

AC AD145176;

DT 22-APR-2004 (first entry)

DE Rice isoprenoid biosynthesis-associated cDNA #54.

Rice; isoprenoid biosynthesis; ss; gene; plant; isopentenyl diphosphate;
IPP; dimethylallyl alcohol; DMAPP; short-chain plastid prenyltransferase;
gibberellin; carotenoid; abscisic acid; tocopherol; plastoquinone;
phyloquinone; mevalonate pathway; phytosterol; brassinosteroid;
haeme; yield.

Oryza sativa.

US2004010815-A1.

15-JAN-2004.

26-SEP-2002; 2002US-00259194.
26-SEP-2001; 2001US-0325277P.
04-APR-2002; 2002US-0370620P.
04-APR-2002; 2002US-0370743P.
(LANG/) LANGE B M.
(GHAS/) GHASSEMIAN M.
(BRIG/) BRIGGS S P.
(COOP/) COOPER B.
(GLAZ/) GLAZEROOK J.
(GOFF/) GOFF S A.
(KATA/) KATAGIRI F.
(KREP/) KREPS J.
(MOUG/) MOUGHAMER T.
(PROV/) PROVART N.
(RICK/) RICHE D.
(ZHUT/) ZHU T.
Lange BM, Ghassenian M, Briggs SP, Cooper B, Glazebrook J;
Goff SA, Katagiri F, Kreps J, Moughamer T, Provart N, Ricke D;
Zhu T;
WPI; 2004-090562/09.
P-PSDB; ADI45177.
New isolated polynucleotides and polypeptides associated with isoprenoid
synthesis in plants, useful for producing transgenic plants, for targeted
gene disruption, as well as markers or probes.
Claim 1; SEQ ID NO 107; 117pp; English.
The invention relates to a polynucleotide (or its complement, protein
encoding fragment or reverse complement), comprising a nucleotide
sequence encoding a polypeptide comprising an amino acid sequence
involved in or associated with the biosynthesis of isoprenoids in a rice
plant. Also included are an isolated polypeptide involved in or
associated with the biosynthesis of isoprenoids in a plant, an expression
cassette comprising the polynucleotide, a host cell comprising the
expression cassette, and a transgenic plant comprising the expression
cassette. The polypeptides and polynucleotides include those associated
with the biosynthesis of isopentenyl diphosphate (IPP) and dimethylallyl
alcohol (DMAPP), the biosynthesis of short-chain plastid
prenyltransferases, the biosynthesis of gibberellins, the biosynthesis of
carotenoids and/or abscisic acids, the biosynthesis of tocopherols,
plastoquinone and/or phyloquinone biosynthesis, the mevalonate pathway,
phytosterol and brassinosteroid metabolism, biosynthesis of ubiquinone,
biosynthesis of chlorophyll or haeme. Also disclosed are banana, wheat
and corn homologues of some of the rice polynucleotides. The
genome is augmented by a nucleic acid molecule of the invention, or in
which the corresponding gene has been disrupted, e.g. to result in a
loss, a decrease or an alteration in the function of the product encoded
by the gene. The plants may also have increased yields and/or produce a
better quality product than the corresponding wild-type plant. The
nucleic acid molecules are useful for targeted gene disruption, as well
as markers and probes. Note: The sequence data for this patent did not
form part of the printed specification, but was obtained in electronic
format directly from USPTO at
seqdata.uspto.gov/sequence.html?docID=20040010815. The present sequence
is a Rice isoprenoid biosynthesis- associated cDNA of the invention.
Sequence 2223 BP; 552 A; 528 C; 634 G; 509 T; 0 U; 0 Other;

Query Match 13.6%; Score 379.4; DB 12; Length 2223;
Best Local Similarity 54.5%; Pred.No. 3.4e-79;
Matches 900; Conservative 0; Mismatches 676; Indels 75; Gaps 4;

Qy 795 GGATACCTGGCGTATATCTCTGAGGAGGACTCGGTAATTTATATGATGATGGAATATGCGTGAAG 854
Db 574 GAATATACGGCTTATGTCTCAGAGGATAGGCAATATTCAGAACTGGAATGAAGTGTATG 633

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents DNA encoding a diterpene synthase polypeptide of the
CC invention. Note: the sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1559 BP; 453 A; 288 C; 362 G; 456 T; 0 U; 0 Other;

Query Match 11.8%; Score 330; DB 12; Length 1559;
Best Local Similarity 55.7%; Pred. No. 1.4e-67;
Matches 655; Conservative 0; Mismatches 515; Indels 6; Gaps 1;

QY 1270 CGTCACATATATATACCAAGAGATTTATCTCTCTGAAACAAATCTTCAAGTCAGCTG 1329
DB 116 CGACAGTTAGTATCTCTGAAGATGAGTCTATCTCTGATAGCATAGGCTCAAGTCCAGTA 175
QY 1330 ATTTCTCTCAAGAGATAATATCCACTGATTCATCAACACAGGCTTTCTAAATTAATTCACAAAG 1389
DB 176 CCTTACTGAGGGAACACTAGAGTCTGCTGCTCTACGAAAAACCTTCACTCTTTAAAG 235
QY 1390 AGTGGAAATATCTTTAAGTTCCTATCAATACCGGTTTAGAACGATTAACATACATAGAC 1449
DB 236 AGGTGGAACATGCTCTGGACGGTCCCTTCTACACCAATGGACCGTCTACACCATAGGT 295
QY 1450 GAAATATACAGCTTTACAAATGTAGACAAATACAAAGAAATCTGAAACCTACATATCACTCAT 1509
DB 296 GGAACATCGAAATTTCAATATATATAGACGACGACATGCTAGACACCATATCTTGCAA 355
QY 1510 CAAATATAGTAACTGATTAACCTAAGGTTGGCTGTGTAAGATTTCTACCTGGCCAAAT 1569
DB 356 ATCAACATACAGTAGAGATATCTAGCGTTGAGTATAGAGACTTCAGTTCTCTCTCAGT 415
QY 1570 CTATTTATCGTGAAGATTAAGGTTCTTGAAGGTTGGGTGAGAGATAAGTTGGACC 1629
DB 416 TTACTTACCAAGCAAGACTTCAACATCTTGAAGCTGGGTGAAAGAGTGAGGTTAGACC 475
QY 1630 AGCTCAAGTTTGTAGCAAAAGACCGCTACTCTATTTCTCTGTGCTGCAACACTTT 1689
DB 476 AGCTACAAATTTGCGGACAGAGTTGGCATATCTTACTTGTCTGCTGGCCACCATGT 535
QY 1690 CGTCTCCGAAATTAATCAGATCGCGTAATTCATCGGCCCAAAATGGGCATATTAATCTACAG 1749
DB 536 TCTCTCTGAGCTGCTGATCTCGAACTTTGTGGGCCAAAATGGTGTGCTCACAACATA 595
QY 1750 TAGTTGATGACTTTTGTGATATCGGTGATCAATCGATGATTCACCAACCTGATTCAT 1809
DB 596 TTGTTGACGACTCTTTGATGTTGCGGATCAAAAGAGAACTTGAACACCTTGTCTAGT 655
QY 1810 GTGTTGAAAATGGAATAGATGTCGACAGAGGATTTGTTGTCAGAGCATGTTCCGATTT 1869
DB 656 TGGTTGAGATGTTGGACGAGCATCACAAAGTTGAGTTCTACTCAGAACAAAGTAGAGATTA 715
QY 1870 TATTTTATGCAATTAAGATGCAATCTGTTGGATTTGGAGTGAAGCTTTAAATGGCAAG 1929

DB 716 TATTTTCTTCAATTTTANGACTCAGTTAACCAACTTGGTGAAAGGCTTCTTTGGTACAAG 775
QY 1930 CGCGGATCTAAGTACCCATGTTATTCAAACTTGGTTGGAACTAATGATAGTATGTTGA 1989
DB 776 ACCGAGTATTAACCAACACCTAGTAGAATATGGTTGGATTGCTTAAAGTCTATGATGA 835
QY 1990 GAGAAGCTATATGGACAAGAGATGCTATGTGTCACCAACATTAATGAATATATGGAACAG 2049
DB 836 CAGAGGTAGAGTGGCGTTTGGACAAATATGTGCCAACAGAGAAGGAATACATGATAAATG 895
QY 2050 CTTACGTGTCATTTGATAGCCCGATTTGCAAGCCGCTATTTTACTTTTGGGGCCCA 2109
DB 896 CATCTTTAACTTTGGACTAGGCCCAATTTGATTTGCCAGCAATATATTTTGGGGCCAA 955
QY 2110 AATTATCAGAGGAGATTTGTTGAAAGCTCTCAATATCATATATCTATTTAAGCTAATGAGCA 2169
DB 956 AGATCTCAGAGTCTATTTGTTAAAGATCCAGATATGATGAATTTGTTCAACTGATGAGCA 1015
QY 2170 CGCAGGTCGACTTTTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGAAGGCAAT 2229
DB 1016 CATGTGCTGCGCTCTTGAATGATGTTTCAGACTTTTTCAGAGGGAGTACACGAGGGCAAGT 1075
QY 2230 TAAACCGGTAGCATTTGCAATTTGAGTAAACGGAAGTGGAAAGTGGAAAGAGAGGTTG 2289
DB 1076 TGAATAGTGTCTCTCTCTGCTTCTTCATGTGCTCCATGTCCATATCAGACGCCAAA 1135
QY 2290 TGGAGGAGATGATGATGATGATTTTAAACCAAGAGGAAAGAAATTAATGAAATTAATTTTG 2349
DB 1136 GGAATTTACAGAAGCCCATAGACATGACAGAGACACCTCTCTAAGTTTGTTCCTGCTG 1195
QY 2350 AAGAAATGATAGCATGTTCTTAGAGCTTTGTAAGATGCAATTTTGGAAACATGTTGTCAG 2409
DB 1196 AAGAAAGTG-----TTGTTCTTAGGCCCTGCAAGGAATTAATTTTGGAAAATGTGCAAGG 1249
QY 2410 TGTGTAATTTTTTTTAGGCAACGATGACGGGTTTA 2445
DB 1250 TGTGCTACTTCTTCTACTCGACGCGGATGGGTTTA 1285

RESULT 14
ADA70068
ID ADA70068 standard; DNA; 1875 BP.
XX
AC ADA70068;
XX AC
XX AC
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3391.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.

xx Claim 6; SEQ ID NO 3391; 899pp; English.
ps
xx
cc The present invention relates to a method (M1) for identifying genes
cc involved in plant resistance or response to pathogenic infection. M1
cc comprises identifying a gene whose expression is significantly altered in
cc the incompatible interaction of plant gene expression relative to
cc expression of the gene in an uninfected plant, in a mutant plant that
cc does not express a gene associated with response to pathogenic infection,
cc or in a corresponding incompatible or compatible interaction. (M1) is
cc useful for conferring resistance to resistance or tolerance to a plant to
cc bacterial, fungal or viral infection. The present sequence was used to
cc illustrate the invention.
xx
SQ Sequence 1875 BP; 483 A; 415 C; 499 G; 477 T; 0 U; 1 Other;

Query Match 11.8%; Score 329.4; DB 8; Length 1875;
Best Local Similarity 52.4%; Pred. No. 2e-67;
Matches 869; Conservative 1; Mismatches 747; Indels 42; Gaps 5;

QY 324 AAACGTTTAAAGATGAGAAATTTCTGTTTCTTCATATGACACAGCAGGCTAGCCATG 383
Db 208 AAGCAGTCGAGGGGTTGAGTTATCACCATTTTCATACGACGCGCATGGGTGGCCATG 267

QY 384 GTCCCTTCTCCAAACTCACCCAAATCGCCTTGTTCCTCGAGTGTCTCAATTTGGTTAAT 443
Db 268 GTGCCAGTCGAGGGGTCGTGAGTCTCCATGCTTCCGCGAGTGTGCTGAGTGGATCTT 327

QY 444 AATAATCAGCTTAATGATGGTTCATGGGTCCTGTTTAATCACACTCAATATCATATCAC 503
Db 328 CAGAACCAACAGGAGGCGGATCATGGG-----CCATTCTGCGGGCCATCAGGA 378

QY 504 CGTGTCTTAAAGATCTCTATCTTCAACATTAGCATGTATTTGTCATTAAGAGATGG 563
Db 379 GAGGTAAACAAGATATTTCTATGTCTACATGGCATGTGTTCTTGCACTCAACATATGG 438

QY 564 AATGTTGGGGAAGATCAATAAATAAGGTCTAAAGTTTATGAGTCAATCTTGCTTCA 623
Db 439 AATGTTGGTCAAGTACATCAGGAGGACTGAGTTTCAATTTGGAGAAATTTTCGGTT 498

QY 624 GTTACTGAAAAGTCAACATCTCCCATGTTGTTGACATCATATTTCTCGTTTGGCTT 683
Db 499 GCCATTGATGGGCAATGTGCGCTCTCTGTCGGTTTAAACATTACATTTTCTGGTATGCTT 558

QY 684 GAGTATGCGAAAACCTTGGACATAACCTTCTTCAACACACAGATTTAGTTTGATG 743
Db 559 CCGCTTGCCTTGGGATGGTTTGAATTTCCCTGTTATGAGAAACGGATATCGATAGTATA 618

QY 744 CTACATAAGAGGAATTTGGAGCAAAAA-----GATGCCATTCAAATGAGATGGAT 794
Db 619 TTTCCGCTCCGGAGGTGCAATTTGAAAGGATGCTGGTGGCACAGCTTCACAAGGAAA 678

QY 795 GATATCTTGGCGTATATCTTGAAGACCTCGGTAATTTATATGATTTGGAATATGGTGAAG 854
Db 679 GCCTTCATGGCATATGATATCAGAAGGTTGGGAGAGAACAAAGATTGGGATCATGTAATG 738

QY 855 AAATATCAGATGAAAATGTTCTGTTTCACTTCCATCAGATCAGACAGCTCTGCTTTC 914
Db 739 GCATATCAAGAGAAAGACGGTTTCATTTGTTCACTCGCTTCCACCAACAGCAGCTCAGG 798

QY 915 ATTAATCATCAAAATCTCGTTGTCTTAATTTAAATTTCACTTTTGGACAAATTTGGT 974
Db 799 ATCCACAGCTCAATGACAGAGCTTAGACTATTTGGTCTCTTACATCCAAATGGGT 858

QY 975 AATGAGTCCCAACAGTTTATCTCATGATTTTATTTATCTCGACTTTCTATGTTGACACA 1034
Db 859 GGGCCAGTGGCCAGCATATWTCCAGACAGGTATATCTCCAGCTCTGCAATGGTGTATACC 918

QY 1035 ATTGAAGATAGGAATTTCAACCATTTTCAGATGGAAATTAATAATGTTTATAGTAA 1094
Db 919 CTTGAGAGATGGGATCTCTTCTGATTTGCTGCGACATACGGGACATATATGACATG 978

QY 1095 ACATACAGATGTTGGGTGGAACGAGATGAGCAATATTTCTATGATGTTGTAACTATGCT 1154

Db 979 ACTACAGTTGCTGGATGCAAGATGAGGAGGAGATCATGTTGGATATGGCAAGCTGTGCA 1038
QY 1155 TTAGCCTTTGGTTATTAAGGATCAATGGGTATGAGTTTCCCGAGATCAATTTGGCTGAA 1214
Db 1039 AAGGCAATTTGGCTCTCTGATGACGCTTATGAGTCAAGTCTCAGAGGAAATGGCTCGA 1098
QY 1215 ATTACTAATGAATTAGCTTTGAAAGACGAAATATGCGAGC-----TCTTGAACA 1262
Db 1099 TTTGCTGAAAGATGATGCTTTGATGATTCATTTCAATTCATGATATCTCAATGACACTAGCC 1158
QY 1263 TATCATGCTGCATATATATTAACCAAGAGGATTTATCTTCTGGAAAACAAATCTTGAAG 1322
Db 1159 TTGCTGGAGTTGTACAAAGTTTCAAGTTCTCTTCTTGGAAAGAGATTTCTACTACTAGAG 1218
QY 1323 TCAGCTGATTTCTCAAGAGATATATCACTGATTTCA-----AACAGGCTTTCT 1373
Db 1219 AATATCGGTTCTGCTGCTAGCTAAATTTACTCAAGCAACAACTCTCGTTCAACAGATATCG 1278
QY 1374 AATTAATTCACAAAGAGGTGAAAATGCTCTTAAGTTTCCCTATCAATACCGGTTTAGAA 1433
Db 1279 AATCATTAAGCCAGAGGTTGAGTATGCCCTTAAATATCCCTTCTATGCAACCGGTGAA 1338
QY 1434 CGCATAAACACACTAGACGAAATATACAGCTTTTCAATGTAGACAAATACAGAAATCTGAAA 1493
Db 1339 GTACTAGACACAGGGGAAACATAGAGCGAATTCAGCGTAAACGGCTTCAGCGTCTGAAA 1398
QY 1494 ACTACATATCACTCATCAAAATATTAGTAACACTGATTTACTTAAGTTGGCTGTGAAAGAT 1553
Db 1399 TCTGGATATCTGGGCTC---CGGGCAGACAAAGAAATTTCTAGCTTTGCTGTCAATAAA 1455
QY 1554 TTCTACACTGCGCAATCTATTTATCGTGAAGAAATTAAGAGTCTTGAAGGTTGGGTGGTA 1613
Db 1456 TTCCACTAGCGCCAGCTGTTTACCAACAAGAACTCCGGTATCTCGAGAGCTGGGTGGCA 1515
QY 1614 GAGAAATAGTTTGGACAGCTCAAGTTTGTAGGCAAAAAGACCGCTACTGTTATTTCTCT 1673
Db 1516 GAATTTAGCTGGACAGCTGAAATTCGGAGAGTATTCACACTGAGTCGCTTTGTCT 1575
QY 1674 GTTGTGCAACACTTTCTGCTCCGAAATTTATCAGATGCGCGTATTTTCATGGGCCAAAAAT 1733
Db 1576 GCTGTTGCTCCCTGTTCTCCCTGCGAGTTGTCGATGCCGCTTGTGGAGCCAGAAC 1635
QY 1734 GCATATTAACCTACAGTATGATGATCTTTTGAATATCGGTGGTACAAATGATGAATG 1793
Db 1636 GCCATCTAGCGGCTGTGTTGATGACTGTTTCGACGGCGGAGGCTCCATGGAGGAGATG 1695
QY 1794 ACCAACCTGATTCATGTTGAAAATATGGAATGTAGATGTGCAACAGGATTTGTTTCA 1853
Db 1696 CTGAATTTGGTCCGCTTATTCGAAATGGGACGATCAGGTGAGATTTGGCTTCTGCTCC 1755
QY 1854 GAGCATTTGCGAATTTTATTTTAGCATTAAGAGTCAATCTGTTGGATTTGGAGATGAA 1913
Db 1756 AGCAACGTAGAGATATGTTTCAATGCTGTTTACACACGACCAAGCGTATCGCGCAAG 1815
QY 1914 GCTTTTAAATGGCAAGCGCGATGTAATAGCCATGTT 1952
Db 1816 GCCGCACTGGTGCAGAGCGTTGTGTCATCGATCATATT 1854

RESULT 15
ADA69972
ID ADA69972 standard; DNA; 2193 BP.
XX
AC ADA69972;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 3295.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX

OS Oryza sativa.
XX WO2003000898-A1.
XX
XX
XX 03-JAN-2003.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX
XX Identifying at least one gene involved in plant resistance or response to
XX PT pathogenic infection for conferring resistance or tolerance to a plant to
XX PT bacterial, fungal or viral infection by determining or detecting plant
XX PT gene expression.
XX
XX Claim 6; SEQ ID NO 3295; 899pp; English.
XX
XX The present invention relates to a method (M1) for identifying genes
XX involved in plant resistance or response to pathogenic infection. M1
XX comprises identifying a gene whose expression is significantly altered in
XX the incompatible interaction of plant gene expression relative to
XX expression of the gene in an uninfected plant, in a mutant plant that
XX does not express a gene associated with response to pathogenic infection,
XX or in a corresponding incompatible or compatible interaction. (M1) is
XX useful for conferring resistance to resistance or tolerance to a plant to
XX bacterial, fungal or viral infection. The present sequence was used to
XX illustrate the invention.
XX
XX Sequence 2193 BP; 593 A; 440 C; 564 G; 596 T; 0 U; 0 Other;
Query Match 11.3%; Score 315.6; DB 8; Length 2193;
Best Local Similarity 58.1%; Pred. No. 3.9e-64;
Matches 615; Conservative 0; Mismatches 434; Indels 9; Gaps 3;
QY 1388 AGAGGTGGAAATGCTTTAAGTTCCTATCAATACGTTTACGCGTTTGAAGCGCAATACACTAG 1447
Db 1044 AAAAATGAAGTATGCTCTTAAGTTCCCTTCTACACCACACTGCGCGCTAGATCAAA 1103
QY 1448 ACGAATATACACTTTACATATGACATACAGAAATTTCTGAAACTACATATCACTC 1507
Db 1104 GAGGAACATTTGAACGTTTTCATGCAAGAGATTCACAGATGTTAAAGACGGAAT---ACTT 1160
QY 1508 ATCAATATTTAGTAACACTGATTAACCTTAAGTTGGCTGTTGAAGATTTCTACACCTGCCA 1567
Db 1161 GCTTCTCTCATGCCAATCAAGATATTTCTAGCTTTGGCTGTTGAAGATTTCTAGTAGTTCTCA 1220
QY 1568 ATCTATTATTCGTGAAGAAATTAAGAGCTTTGAAGGTGGGTGGGTGAGAAATTAAGTTGGA 1627
Db 1221 ATCTATATACAGGATGAATTAATTAATCTTGAAGTGGGTGAAAGATGAAGAGCTCGA 1280
QY 1628 CCGAGCTCAAGTTTGTAGGCAAGACCGCTACTGTTATTTCTCTGTTGTCGCAACACT 1687
Db 1281 TCAGCTGCCAATTTGCAGCCAGAGTTGACATATTTGCTTTCTGCTGCTGCTACAT 1340
QY 1688 TTGCTCTCCGGAATFATCAGATCGCGTATTTTCATGGGCCAAATAATGGCATATTAATAC 1747
Db 1341 ATTCCCCCGTGAATTTGCTGAAGCCCGCAATTCATGGGCTTAAATAATGGTGTACTGACAAC 1400
QY 1748 AGTAGTTGATGACATTTTGTATATCGGTGGTACATCGATGAATTTGACCACTGATCA 1807
Db 1401 TGTGTTGATGACTTCTTTGATCTTGGGGATCAAAAGAGAACTAGAAAACCTCATTCG 1460
QY 1808 ATGTGTTGAAAATGGAATGAGATGTCGACAGGATTTGTTTCAGAGATGTTGCGAT 1867
Db 1461 TTTAGTTGAGAAGTGGGATGGACAT---CAGAGAGTCTTACTCAGAACAGTAAGAAAT 1517

QY 1868 TTTATTTTGTAGCATTAAAGATGCAATCTGTTGGATTGGAGATGAAGCTTTTAAATGGCA 1927
Db 1518 AGTTTTTCTGCTATTTATATACAGTGAACAGCTTGGAGCAAGGCTTCTGCATTACA 1577
QY 1928 AGCGCCGATGTAACTAGCCATGTTATTCAAACTTGGTTGGAACTAATGAATAGTATGTT 1987
Db 1578 AGCCCGTGTATTTACAAAACACCTTAACAGAAATATGTTATGTTCTGATGAGGCTATGAT 1637
QY 1988 GAGAGAAGCTATATGACAAAGAGATGCTTATGTCACCAACATTAATTAATATATATGAAAA 2047
Db 1638 GACCGAAGCTGAATGGCAGAGGACAAAATATGTCGCACAATGGAAGATATATGCAAA 1697
QY 2048 CGCTTACGTGCTCATTTGCAATTAGCCCGGATTTCAAGCCGGCTATTACTTTTGGGGCC 2107
Db 1698 TGCTGTTGCTCATTTTGCACCTGCGGACCTATTGCTCCCGACTCTGTATTTCTGTAGGACC 1757
QY 2108 CAATATATCAGAGAGATTTGTTGAAAGCTCTGATATATCAATCTATTTAAGCTAATGAG 2167
Db 1758 GAAGCTCCAAGAGGATGTCGTACAGGATCATGAGTACAATGAATTAATTTAGACTGATGAG 1817
QY 2168 CACGACGGGTGCACTTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTTAAGGAAGGCAA 2227
Db 1818 CACTTGTGGCGCTCTCTGAAATGACAGCCAGGCTTTGAGAGGAGAGCCCTGGAGGAAA 1877
QY 2228 ATTAAGCGCGGTAGCATTTGATTAACGAGAAAGTGGAAAGTGGAAAGTGAAGAGGT 2287
Db 1878 GCTGAACAGTGTCTCACTGCTTCTTCTATCAGCTGCTGTTCTATCTCCATAGACGAGGC 1937
QY 2288 TGTGGAGGAGATGATGATGATTAATAAACAAGAGGAAAGATTAATGAATTAATTTT 2347
Db 1938 TAAATGAAGGCCAGAAATCCATAGACACTTCCAGGAGAACTTGTAAAGATTGGTCCCT 1997
QY 2348 TGAAGAAAATGGTGTAGCATTTGTTCTAGAGCTTTGTAAGATGCAATTTTGGAACTGTGCA 2407
Db 1998 TGAAGAACAGGT---GCTGTTCTTAGGCCATGCAAGCAGCTGTTCTGGAAGATGTCAA 2054
QY 2408 CGTGTGAATTTTTTTTACGCAAAACGATGACGGTTTA 2445
Db 2055 GATTGTTCAATGTTTCTACTCTCGGACTGATGGGTTTA 2092

Search completed: October 16, 2004, 23:38:25
Job time : 1312.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 15:57:41 ; Search time 11750.6 Seconds
(without alignments)
11236.263 Million cell updates/sec

Title: US-10-041-018-361
Perfect score: 2792
Sequence: 1 cccactcatcctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1 | 2792 | 100.0 | 2792 | 8 | AF097311 | AF097311 Stevia re |
| 2 | 2582.6 | 92.5 | 3117 | 8 | AF097310 | AF097310 Stevia re |
| 3 | 1428.6 | 51.2 | 2594 | 8 | AB031205 | AB031205 Lactuca s |
| 4 | 809.4 | 29.0 | 2450 | 8 | AB045310 | AB045310 Cucumis s |
| 5 | 782.2 | 28.0 | 2858 | 6 | E12936 | E12936 cDNA encodi |
| 6 | 782.2 | 28.0 | 2858 | 8 | CM043904 | U43904 Cucurbita m |
| 7 | 732.4 | 26.2 | 2506 | 8 | AF034774 | AF034774 Arabidops |
| 8 | 559.8 | 20.1 | 2271 | 8 | AY347876 | AY347876 Oryza sat |
| 9 | 555 | 19.9 | 2283 | 8 | AY347877 | AY347877 Oryza sat |
| 10 | 543.4 | 19.5 | 2283 | 8 | AY347878 | AY347878 Oryza sat |
| 11 | 530.8 | 19.0 | 3234 | 8 | AK119442 | AK119442 Oryza sat |
| 12 | 479.6 | 17.2 | 2529 | 8 | AY347880 | AY347880 Oryza sat |
| 13 | 467.6 | 16.7 | 2753 | 8 | AK119327 | AK119327 Oryza sat |
| 14 | 453.2 | 16.2 | 2403 | 6 | AR452826 | AR452826 Sequence |
| 15 | 448.2 | 16.1 | 3061 | 8 | AK068310 | AK068310 Oryza sat |
| 16 | 444.6 | 15.9 | 2772 | 8 | AB089272 | AB089272 Oryza sat |
| 17 | 434.8 | 15.6 | 2497 | 8 | AY347882 | AY347882 Oryza sat |
| 18 | 416.8 | 14.9 | 2154 | 6 | AX653583 | AX653583 Sequence |
| 19 | 411.4 | 14.7 | 2888 | 8 | AK121446 | AK121446 Oryza sat |

| | | | | | | |
|----|-------|------|------|---|----------|--------------------|
| 20 | 386.4 | 13.8 | 2813 | 8 | AK072461 | AK072461 Oryza sat |
| 21 | 379.4 | 13.6 | 2223 | 6 | AX653424 | AX653424 Sequence |
| 22 | 378 | 13.5 | 2220 | 8 | AY347879 | AY347879 Oryza sat |
| 23 | 368.6 | 13.2 | 2463 | 8 | AB118056 | AB118056 Oryza sat |
| 24 | 355 | 12.7 | 1773 | 8 | AF529266 | AF529266 Zea mays |
| 25 | 332.4 | 11.9 | 2149 | 8 | AY347881 | AY347881 Oryza sat |
| 26 | 330 | 11.8 | 1559 | 8 | AF105149 | AF105149 Zea mays |
| 27 | 329.4 | 11.8 | 1875 | 6 | AX653521 | AX653521 Sequence |
| 28 | 315.6 | 11.3 | 2193 | 6 | AX653425 | AX653425 Sequence |
| 29 | 241.6 | 8.7 | 2313 | 6 | AX653751 | AX653751 Sequence |
| 30 | 238.2 | 8.5 | 1157 | 6 | AR452831 | AR452831 Sequence |
| 31 | 194.8 | 7.0 | 2861 | 6 | BD227676 | BD227676 Synthesas |
| 32 | 194.8 | 7.0 | 2861 | 6 | AR240711 | AR240711 Sequence |
| 33 | 194.8 | 7.0 | 2861 | 6 | AR266986 | AR266986 Sequence |
| 34 | 194.8 | 7.0 | 2861 | 6 | AR316335 | AR316335 Sequence |
| 35 | 194.8 | 7.0 | 2861 | 6 | AR338478 | AR338478 Sequence |
| 36 | 194.8 | 7.0 | 2861 | 6 | AR429884 | AR429884 Sequence |
| 37 | 194.8 | 7.0 | 2861 | 6 | AGU50768 | U50768 Abies grand |
| 38 | 166.6 | 6.0 | 2643 | 8 | AY424738 | AY424738 Taxus bac |
| 39 | 166.6 | 6.0 | 2771 | 8 | AY461450 | AY461450 Taxus x m |
| 40 | 160.2 | 5.7 | 1568 | 8 | AK108710 | AK108710 Oryza sat |
| 41 | 158.6 | 5.7 | 2706 | 8 | AY007207 | AY007207 Taxus chi |
| 42 | 153.8 | 5.5 | 2700 | 6 | AR110171 | AR110171 Sequence |
| 43 | 153.8 | 5.5 | 2700 | 6 | BD227670 | BD227670 Synthesas |
| 44 | 153.8 | 5.5 | 2700 | 6 | AR240705 | AR240705 Sequence |
| 45 | 153.8 | 5.5 | 2700 | 6 | AR266980 | AR266980 Sequence |

ALIGNMENTS

| | | | | | | |
|------------|--|-------------------|---------------------------|------|---------------|-----------------|
| RESULT 1 | AF097311 | AF097311 | 2792 bp | mRNA | linear | PLN 22-MAR-2000 |
| LOCUS | Stevia rebaudiana | Stevia rebaudiana | kaurene synthase (KS22-1) | mRNA | complete cds. | |
| DEFINITION | AF097311 | AF097311 | GI:4959240 | | | |
| ACCESSION | AF097311.1 | AF097311.1 | GI:4959240 | | | |
| VERSION | AF097311.1 | AF097311.1 | GI:4959240 | | | |
| KEYWORDS | Stevia rebaudiana | Stevia rebaudiana | | | | |
| SOURCE | Stevia rebaudiana | Stevia rebaudiana | | | | |
| ORGANISM | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Asteroidae; Eupatorieae; Stevia. | | | | | |
| REFERENCE | 1 (bases 1 to 2792) | | | | | |
| AUTHORS | Richman,A.S., Giljen,M., Starratt,A.N., Yang,Z. and Brandlie,J.E. | | | | | |
| TITLE | Diterpene synthesis in Stevia rebaudiana: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway | | | | | |
| JOURNAL | Plant J. 19 (4), 411-421 (1999) | | | | | |
| MEDLINE | 99435880 | | | | | |
| PUBMED | 10504563 | | | | | |
| REFERENCE | 2 (bases 1 to 2792) | | | | | |
| AUTHORS | Richman,A.S., Giljen,M. and Brandlie,J.E. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (06-OCT-1998) Southern Crop Protection and Food Research Center, Agriculture and Agri-Food Canada, 1391 Sandford Street, London, ON N5V 4T3, Canada | | | | | |
| FEATURES | Location/Qualifiers | | | | | |
| source | 1..2792 | | | | | |
| | /organism="Stevia rebaudiana" | | | | | |
| | /mol_type="mRNA" | | | | | |
| | /db_xref="taxon:55670" | | | | | |
| | /tissue_type="leaf" | | | | | |
| | 1..2792 | | | | | |
| | /gene="KS22-1" | | | | | |
| | 174..2528 | | | | | |
| | /gene="KS22-1" | | | | | |
| | /function="synthesis of (-)-kaurene by ionization dependant cyclization of copalyl diphosphate" | | | | | |
| | /note="terpene cyclase" | | | | | |
| | /codon_start=1 | | | | | |
| | /product="kaurene synthase" | | | | | |
| gene | | | | | | |
| CDS | | | | | | |

/protein_id="A034295.1"
/db_xref="GI:4959241"
/translations="MNLISCIASPLLTLSNRPAALSAIHTASTSHGGQNTPTNLIIDT
TKERIQKFNKIVISYSYDTAWVAMVSPNSPKSPFPECLANLNNQINDGSGWLV
NHTNNHNPDLKDSLSLACI VALKRWNVGDEQINKGLSFISNLSASATEKQSPSI
GPDILPGLLEYAKNLDINLSKQDPSLMHLKRELBOKRCHSNEMDGYLAYSIGLG
NLYDMNVKVKQKNGSVFNSPATAAFINHONPGLNLYLSLIDKFGNAVPTVPH
DLFIRLSNWDITIELGISHFRVEIKNVLDETRCWVERDEQIFMDVVTCAALFRLIR
INGVEVSPDPLAETINELALKDEYALETHASHLIDYOEDELSKQLKSDAFELKEI
STDSNRLSLIHKREVENALFPINTGLERINRNRNIQIYNVDNTRILKTYHSSNTSN
TDYLRILAVEDFYTCOSIYREBELKGLRWVVENKLDQKFARQKATCYFVSATLSSP
ELSDARISWAKNGILTTVDPDFDIDGTDIDELTNLIOCVEKNVDVDCSCSEHVRIL
FLAKDAICWIGDEARFQWQARDVTSHVIOIWLLELMSLREAIWTRDAYVPTLNEYME
NAYVSFALGIVKPAIYFVGPKLSEIEVSESEYHNLFLKMSLQGRLLNDLHSPKREPK
EKLNALVHLHNSGESKEVEEVVMMKIKRKLMLKLIIFRENGSIVPRACKDAF
WNWCHVLNFFYANDDGTGNTILDITVDKIIYNPLIVNVNEBQR"

ORIGIN

| | | Query Match | 100.0%; | Score 2792; | DB 8; | Length 2792; |
|----|-----|---|-----------------|---------------|-----------|--------------|
| | | Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| | | Matches 2792; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | CCCACTCATCCTTTATCAACCAATACCATCGTTCGCCACCGGAAGACTGATACGCGACG | 60 | | | |
| Db | 1 | CCCACTCATCCTTTATCAACCAATACCATCGTTCGCCACCGGAAGACTGATACGCGACG | 60 | | | |
| Qy | 61 | GACCTGAGCAGCGCTTTAATCTCTGTGTGGACCACTAACAAAAATTCGACCAATCAGATCT | 120 | | | |
| Db | 61 | GACCTGAGCAGCGCTTTAATCTCTGTGTGGACCACTAACAAAAATTCGACCAATCAGATCT | 120 | | | |
| Qy | 121 | ACTCCGGTGGACAGTTTTCAATTTGCAATTTAGTAATCAGCAAACTAAACATCAATGAATC | 180 | | | |
| Db | 121 | ACTCCGGTGGACAGTTTTCAATTTGCAATTTAGTAATCAGCAAACTAAACATCAATGAATC | 180 | | | |
| Qy | 181 | TTTCACTATGATCGCGTCCCGTTTGTAAACCAATCAAAATCGACCGCGGCTCTGTGAG | 240 | | | |
| Db | 181 | TTTCACTATGATCGCGTCCCGTTTGTAAACCAATCAAAATCGACCGCGGCTCTGTGAG | 240 | | | |
| Qy | 241 | CTATTCATACAGCATCAACTTCACATGSGTGACAACTAATCCCACTAATCTGATCATATG | 300 | | | |
| Db | 241 | CTATTCATACAGCATCAACTTCACATGSGTGACAACTAATCCCACTAATCTGATCATATG | 300 | | | |
| Qy | 301 | ATACAAACAAAGACGGATCAAAAACAGTTTAAATAATGTAGAAATTTCTGTTCTTCAT | 360 | | | |
| Db | 301 | ATACAAACAAAGACGGATCAAAAACAGTTTAAATAATGTAGAAATTTCTGTTCTTCAT | 360 | | | |
| Qy | 361 | ATGACACAGCATGGGTAGCCATGGTCCCTTCACAACTACCCAAATCGCCTGTTTCC | 420 | | | |
| Db | 361 | ATGACACAGCATGGGTAGCCATGGTCCCTTCACAACTACCCAAATCGCCTGTTTCC | 420 | | | |
| Qy | 421 | CTGAGTGTCTCAATTTGGTTTAATTAATACAGCTTAATGATGGTTCATGGGCTCTGTGTA | 480 | | | |
| Db | 421 | CTGAGTGTCTCAATTTGGTTTAATTAATACAGCTTAATGATGGTTCATGGGCTCTGTGTA | 480 | | | |
| Qy | 481 | ATCACACTCATATAATCAATCAACCGGTTGCTTAAAGATTCCTATCTTCAACATTAGCAT | 540 | | | |
| Db | 481 | ATCACACTCATATAATCAATCAACCGGTTGCTTAAAGATTCCTATCTTCAACATTAGCAT | 540 | | | |
| Qy | 541 | GTATTTGGTCATTTAAAGATGGAATGTTGGGGAGAGATCAATAAATAAAGGTCTAAGTT | 600 | | | |
| Db | 541 | GTATTTGGTCATTTAAAGATGGAATGTTGGGGAGAGATCAATAAATAAAGGTCTAAGTT | 600 | | | |
| Qy | 601 | TTATTTGAGTCAAACTCTTGCTTCACTACTGAAAAAAGTCAACCATCTCCCATTTGGTTTG | 660 | | | |
| Db | 601 | TTATTTGAGTCAAACTCTTGCTTCACTACTGAAAAAAGTCAACCATCTCCCATTTGGTTTG | 660 | | | |
| Qy | 661 | ACATCATATTTCTGGTTTGGTTGAGTATGCGAAAAAATTGGACATAAAACCTCTCTTCAA | 720 | | | |
| Db | 661 | ACATCATATTTCTGGTTTGGTTGAGTATGCGAAAAAATTGGACATAAAACCTCTCTTCAA | 720 | | | |
| Qy | 721 | AACAAACAGATTTTGGTTGATGCTACATAAGAGGGAATTTGGAGCAAAAAGATGCCATT | 780 | | | |
| Db | 721 | AACAAACAGATTTTGGTTGATGCTACATAAGAGGGAATTTGGAGCAAAAAGATGCCATT | 780 | | | |

| | | | |
|----|------|---|------|
| Qy | 781 | CAAAATGAGATGGATGGATACCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATT | 840 |
| Db | 781 | CAAAATGAGATGGATGGATACCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATT | 840 |
| Qy | 841 | GGAAATATGGTGAAGAAATATCAGATGAAAAATGGTCTCTGTTTCAAATCCACATCAGCAA | 900 |
| Db | 841 | GGAAATATGGTGAAGAAATATCAGATGAAAAATGGTCTCTGTTTCAAATCCACATCAGCAA | 900 |
| Qy | 901 | CAGCTGCTGCTTTCTTAATTAATCATCAAAATCCTGGTGTCTTAATTTAAATTCACATTT | 960 |
| Db | 901 | CAGCTGCTGCTTTCTTAATTAATCATCAAAATCCTGGTGTCTTAATTTAAATTCACATTT | 960 |
| Qy | 961 | TGGACAGATTTGGTAAATGCGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTT | 1020 |
| Db | 961 | TGGACAGATTTGGTAAATGCGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTT | 1020 |
| Qy | 1021 | CTATGGTTGACAAATTTGAAAGATTTAGGAATTTTACACCAATTTCCAGAGTGAATTTAAAA | 1080 |
| Db | 1021 | CTATGGTTGACAAATTTGAAAGATTTAGGAATTTTACACCAATTTCCAGAGTGAATTTAAAA | 1080 |
| Qy | 1081 | ATGTTTATGATGAAACATACAGATGTTGGGTGGACGAGATGACCAATATTCATGATG | 1140 |
| Db | 1081 | ATGTTTATGATGAAACATACAGATGTTGGGTGGACGAGATGACCAATATTCATGATG | 1140 |
| Qy | 1141 | TTGTAAACATGCTGCTTTTAGCCTTTCCGTTATTAAGGATCAATGGGTATGAAGTTTCCCGAG | 1200 |
| Db | 1141 | TTGTAAACATGCTGCTTTTAGCCTTTCCGTTATTAAGGATCAATGGGTATGAAGTTTCCCGAG | 1200 |
| Qy | 1201 | ATCCATTTGGCTGAAATTAATAAGATTTAGCTTTTGAAGACGAATATTCAGCTCTTGA | 1260 |
| Db | 1201 | ATCCATTTGGCTGAAATTAATAAGATTTAGCTTTTGAAGACGAATATTCAGCTCTTGA | 1260 |
| Qy | 1261 | CATATCATGCTGTCACATATATTAACCAAGAGGATTTATCTCTCGGAAAAACAAATCTTGA | 1320 |
| Db | 1261 | CATATCATGCTGTCACATATATTAACCAAGAGGATTTATCTCTCGGAAAAACAAATCTTGA | 1320 |
| Qy | 1321 | AGTCAGCTGATTTCTCTCAAGAGATAATATCCACTGATTTCAACAGGCTTTCTAAATTTAA | 1380 |
| Db | 1321 | AGTCAGCTGATTTCTCTCAAGAGATAATATCCACTGATTTCAACAGGCTTTCTAAATTTAA | 1380 |
| Qy | 1381 | TTCAAAAGAGGTGGAAAAATGCTCTTAAGTTCCCTATCAATACCGGTTTGAAGCGCATAA | 1440 |
| Db | 1381 | TTCAAAAGAGGTGGAAAAATGCTCTTAAGTTCCCTATCAATACCGGTTTGAAGCGCATAA | 1440 |
| Qy | 1441 | ACACTAGACGAAATATACAGCTTTTACAAATGAGACAAATACAAAGAAATCTGAAAACTACAT | 1500 |
| Db | 1441 | ACACTAGACGAAATATACAGCTTTTACAAATGAGACAAATACAAAGAAATCTGAAAACTACAT | 1500 |
| Qy | 1501 | ATCACTCATCAAAATATTAGTAACACTGATTAACCTAAGGTTTGGCTGTTGAAGATTTCTACA | 1560 |
| Db | 1501 | ATCACTCATCAAAATATTAGTAACACTGATTAACCTAAGGTTTGGCTGTTGAAGATTTCTACA | 1560 |
| Qy | 1561 | CCTGCCAATCTATTATTCGTGAAGAAATTAAGGCTTTGAAGGTGGTGGGTGAGAGAATA | 1620 |
| Db | 1561 | CCTGCCAATCTATTATTCGTGAAGAAATTAAGGCTTTGAAGGTGGTGGGTGAGAGAATA | 1620 |
| Qy | 1621 | AGTTGGACCACTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTG | 1680 |
| Db | 1621 | AGTTGGACCACTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTG | 1680 |
| Qy | 1681 | CAACACTTTCTCTCCGAAATTAATCAGATGCGGTATTTTCATGGGCCAAAAATGGCATAT | 1740 |
| Db | 1681 | CAACACTTTCTCTCCGAAATTAATCAGATGCGGTATTTTCATGGGCCAAAAATGGCATAT | 1740 |
| Qy | 1741 | TAACTACAGTATGATGATCTTTTGTATATCGGTGGTACAAATCGATGAATTTGACCAACC | 1800 |
| Db | 1741 | TAACTACAGTATGATGATCTTTTGTATATCGGTGGTACAAATCGATGAATTTGACCAACC | 1800 |
| Qy | 1801 | TGATTTCAATGTTGAAAAATGGAATGTAGATGTGCAACAGGATTTGTTTTCAGAGCATG | 1860 |
| Db | 1801 | TGATTTCAATGTTGAAAAATGGAATGTAGATGTGCAACAGGATTTGTTTTCAGAGCATG | 1860 |

Db 706 CATGGTCCCTTCCCAACTCACCAATCGCCCTTGTTCCCTGAGTGTCTCAATTGGTT 765
Qy 440 AATTAATAATCAGCTTAATGANGTTTCATGGGGTCTTGTTAATCACACTCATTAATCATAA 499
Db 766 AATTAATAATCAGCTTAATGANGTTTCATGGGGTCTTGTTAATCACACTCATTAATCATAA 825
Qy 500 TCACCCGTTGCTTAAGATCTCTCATCTCAACATTAGCATGTATTGTTGCAATTAATAAG 559
Db 826 TCACCCGTTGCTTAAGATCTCTCATCTCAACATTAGCATGTATTGTTGCAATTAATAAG 885
Qy 560 ATGGAATGTTGGGGAAGATCAATAATAAAGTCTAAGTCTTAAGTCTTAAGTCTTAAGTCTTAAG 619
Db 886 ATGGAATGTTGGGGAAGATCAATAATAAAGTCTAAGTCTTAAGTCTTAAGTCTTAAGTCTTAAG 945
Qy 620 TTGAGTATCGAATAAAGTCAACCATCTCCCATTTGGTCTTGATCATATTTCCCTGGTTT 679
Db 946 TTGAGTATCGAATAAAGTCAACCATCTCCCATTTGGTCTTGATCATATTTCCCTGGTTT 1005
Qy 680 GCTTGAGTATCGAATAAAGTCTTGAGCATAACTCTCTTTCAAAAACAAACAGATTTTAGTTT 739
Db 1006 GCTTGAGTATCGAATAAAGTCTTGAGCATAACTCTCTTTCAAAAACAAACAGATTTTAGTTT 1065
Qy 740 GATGCTACATAAGAGGAAATGGAGCAAAAAGATGCCATTCAAAATGAGATGGATGGATA 799
Db 1066 GATGCTACATAAGAGGAAATGGAGCAAAAAGATGCCATTCAAAATGAGATGGATGGATA 1125
Qy 800 CTTGGCGTATATCTCTGAGGACTCGGTAAATTTATATGATGGAAATATGGTGAAGAAATA 859
Db 1126 CTTGGCGTATATCTCTGAGGACTCGGTAAATTTATATGATGGAAATATGGTGAAGAAATA 1185
Qy 860 TCAGATCAAAAATGCTGTTCTGTTTCACTCAACATCAGCAAGCTGCTGCTTTCAATTA 919
Db 1186 TCAGATCAAAAATGCTGTTCTGTTTCACTCAACATCAGCAAGCTGCTGCTTTCAATTA 1245
Qy 920 TCATCAAAAATCCTGGTGTCTTAATTTAATTTCACTTTGAGCAAGTTTGGTAATGC 979
Db 1246 TCATCAAAAATCCTGGTGTCTTAATTTAATTTCACTTTGAGCAAGTTTGGTAATGC 1305
Qy 980 AGTCCCAACAGTTTATCCTCATGATTTATTTATCGGACTTCTGATGTTGACACAATTA 1039
Db 1306 AGTCCCAACAGTTTATCCTCTTGATTTATATATCCGGCTTCTATGTTGACACAATTA 1365
Qy 1040 AAGATTAGCAATTTACACCACTTTGAGAGTGGAAATTAATAATGTTTATAGATGAACAATA 1099
Db 1366 AAGATTAGCAATTTACACCACTTTGAGAGTGGAAATTAATAATGTTTATAGATGAACAATA 1425
Qy 1100 CAGATGTTGGTGGAAACAGATGAGCAAAATTTATGAGATGTTGTAACATGCTGTTTAC 1159
Db 1426 CAGATGTTGGTGGAAACAGATGAGCAAAATTTATGAGATGTTGTAACATGCTGTTTAC 1485
Qy 1160 CTTTCCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATGCGCTGAAATTA 1219
Db 1486 CTTTCCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATGCGCTGAAATTA 1545
Qy 1220 TAATGAATTTAGCTTTGAAAGCAATATGCAAGTCTTTGAAACATATCATGCGTCAACAT 1279
Db 1546 TAATGAATTTAGCTTTGAAAGCAATATGCAAGTCTTTGAAACATATCATGCGTCAACAT 1605
Qy 1280 ATTATACCAAGGATTTATCTCTGGAATAAATAATCTGGAATCATGCTGATTTCTCTCAA 1339
Db 1606 ATTATACCAAGGATTTATCTCTGGAATAAATAATCTGGAATCATGCTGATTTCTCTCAA 1665
Qy 1340 AGAGATATATCCATGATTCAAACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAAAA 1399
Db 1666 AGGATATATTCCTACTGATTCAAACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAAAA 1725
Qy 1400 TGCTCTTAAGTTCCCTATCAATACCGGTTTGAACGCAATAAACAATAGACGAAATATACA 1459
Db 1726 TGCTCTTAAGTTCCCTATCAATACCGGTTTGAACGCAATAAACAATAGACGAAATATACA 1785
Qy 1460 GCTTTACATGTAGACAAATCAAGAAATCTGAAACTACATATCTCATCAATATTAG 1519

Db 1786 GCTTTACAATGTAGACAAATCAAGAAATTTCTGAAAACTACATATCACTCATCAATATTAG 1845
Qy 1520 TAACACTGATTAACCTTAAGGTTGGCTGTTGAAGATTTCTACACCTGCGCAATCTATTATTCG 1579
Db 1846 TAACACTGATTAACCTTAAGGTTGGCTGTTGAAGATTTCTACACCTGCGCAATCTATTATTCG 1905
Qy 1580 TGAAGAATTAAGAGGCTTTGAAAGGTTGGTGTAGAGAAATAAGTTGACAGCTCAAGTT 1639
Db 1906 TGAAGAATTAAGAGGCTTTGAAAGGTTGGTGTAGAGAAATAAGTTGACAGCTCAAGTT 1965
Qy 1640 TGCTAGCAAAAAGACCGCTTACTGTTATTTCTGTTGTCGCAACACTTTGCTGCTCCGA 1699
Db 1966 TGCTAGCAAAAAGACCGCTTACTGTTATTTCTGTTGTCGCAACACTTTGCTGCTCCGA 2025
Qy 1700 ATTATCAGATGCGGTATTTTCATGGGCAAAAATGCGATATTAACACTAGTAGTTGATGA 1759
Db 2026 ATTATCAGATGCGGTATTTTCATGGGCAAAAATGCGATATTAACACTAGTAGTTGATGA 2085
Qy 1760 CTTTTTTGATATCGGTGGTACAATCGATGAATTGACCAACTGATTCATGTTGTTGAAAA 1819
Db 2086 CTTTTTTGATATCGGTGGTACAATCGATGAATTGACCAACTGATTCATGTTGTTGAAAA 2145
Qy 1820 ATGGAATGTAGATGTGCAAGGATTTGTTGTTGAGCATGTTCCGATTTTATTTTAGC 1879
Db 2146 ATGGAATGTAGATGTGCAAGGATTTGTTGTTGAGCATGTTCCGATTTTATTTTAGC 2205
Qy 1880 ATTAAGAGATGCAATCTGTTGGATGGAGATGAAGCTTTTTAAATGGCAAGCGCGCATGT 1939
Db 2206 ATTAAGAGATGCAATCTGTTGGATGGAGATGAAGCTTTTTAAATGGCAAGCGCGCATGT 2265
Qy 1940 AACTAGCCATGTTAATCAAACTTGTTGGAACTAATGAATGATGTTGAGAGAGCTAT 1999
Db 2266 AACTAGCCATGTTAATCAAACTTGTTGGAACTAATGAATGATGTTGAGAGAGCTAT 2325
Qy 2000 ATGCAAGAGATGCTTATGTCGCAACATTAATAATGAATATATGAAACGCTTACGCTC 2059
Db 2326 ATGCAAGAGATGCTTATGTCGCAACATTAATAATGAATATATGAAACGCTTACGCTC 2385
Qy 2060 ATTTGCAATTAGCCCGGATGTCAGCGGCTTATTTACTTTTGGGGGCCCAATTTATCAGA 2119
Db 2386 ATTTGCAATTAGCCCGGATGTCAGCGGCTTATTTACTTTTGGGGGCCCAATTTATCAGA 2445
Qy 2120 GGAGATGTTGAAAGCTCTGAATATCATATCTAATTTAAGCTAATGAGCAGCGAGGTCG 2179
Db 2446 GGAGATGTTGAAAGCTCTGAATATCATATCTAATTTAAGCTAATGAGCAGCGAGGTCG 2505
Qy 2180 ACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGCAAGCAAAATTAACCGGCT 2239
Db 2506 ACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGCAAGCAAAATTAACCGGCT 2565
Qy 2240 AGCAATTCATTTGAGTAACCGAGAAAGTGGGAAAGTGGAAAGAGGTTGTGGAGGAGAT 2299
Db 2566 AGCAATTCATTTGAGTAACCGAGAAAGTGGGAAAGTGGAAAGAGGTTGTGGAGGAGAT 2625
Qy 2300 GATGATGATGATTAATAAACAAGAGAAAGATTAATGAATTAATTTTGAAGAAATGG 2359
Db 2626 GATGATGATGATTAATAAACAAGAGAAAGATTAATGAATTAATTTTGAAGAAATGG 2685
Qy 2360 TAGCATTTGTTCTTAGAGCTTGAAGATGCAATTTTGGAAACATGTCACGTTTGAATTT 2419
Db 2686 TAGCATTTGTTCTTAGAGCTTGAAGATGCAATTTTGGAAACATGTCACGTTTGAATTT 2745
Qy 2420 TTTTATAACCAACGATGACGGTTTACTGGAAACACGATTTCTTGATATCTGTGAAGGACAT 2479
Db 2746 TTTTATAACCAACGATGACGGTTTACTGGAAACACGATTTCTTGATATCTGTGAAGGACAT 2805
Qy 2480 CATTTACAAACCGTTGCTGTTGATGAATGAAGAAAGAAAGTAATTTGGCTTTAA 2539
Db 2806 CATTTACAAACCGTTGCTGTTGATGAATGAAGAAAGAAAGTAATTTGGCTTTAA 2865
Qy 2540 AAGATGATTAATAATCATGCTCTTGACGGGCTTACTGTTAGTTGTATTAATAAGGTT 2599
Db 2866 AAGATGATTAATAATCATGCTCTTGACGGGCTTACTGTTAGTTGTATTAATAAGGTT 2925

```
QY 2600 GTAGTTGTATATAAAGGTAATAGGTAATCAATAGAGAGCTTAAGTTATTAACTTTTTC 2659
Db 2926 GTAGTTGTATATAAAGGTAATAGGTAATCAATAGAGAGCTTAAGTTATTAACTTTTTC 2985
QY 2660 CTCGTGTCAACACCGGTGAGGTTCTTTGTTAAAGCAGCTTTATTATTATGTTTGTAGTACGACAC 2719
Db 2986 CTCGTGTCAACACCGGTGAGGTTCTTTGTTAAAGCAGCTTTATTATTATGTTTGTAGTACGACAC 3045
QY 2720 TTGGTGGTTGTTGGTTGTTTATTCTCTAAGAGCGTTTCATTGTTCTTAAAAAAA 2776
Db 3046 TTGGTGGTTGTTGGTTGTTTATTCTCTAAGAGCGTTTCATTGTTCTTAAAGAA 3102

RESULT 3
LOCUS AB031205
DEFINITION Lactuca sativa Lsks1 mRNA for ent-kaurene synthase N01, complete cds.
ACCESSION AB031205
VERSION AB031205.1 GI:9971224
KEYWORDS ent-kaurene synthase N01.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
1 (bases 1 to 2594)
Toyomasa, T., Mitsuhashi, W. and Kamiya, Y.
Gibberellin biosynthetic enzyme
Published Only in DataBase (2000)
2 (bases 1 to 2594)
Toyomasa, T., Mitsuhashi, W. and Kamiya, Y.
Direct Submission
Submitted (17-AUG-1999) Tomonobu Toyomasa, Yamagata University,
Faculty of Agriculture; Wakaba-cho 1-23, Tsuruoka, Yamagata
997-8555, Japan (E-mail: toyomasa@tds1.tr.yamagata-u.ac.jp,
Tel:81-235-28-2861, Fax:81-235-28-2812)

FEATURES
Location/Qualifiers
1..2594
/organism="Lactuca sativa"
/mol_type="mRNA"
/strain="Grand Rapids"
/db_xref="taxon:4236"
1..2594
/gene="LsKS1"
87..2453
/gene="LsKS1"
/codon_start=1
/product="ent-kaurene synthase N01"
/protein_id="BAB12441.1"
/db_xref="GI:9971225"
/translation="MNIAQITSSAMLVPSHSHIPHSWVNCMVNPSPGLRTASSQA
GVNPVTMTLVTKRIKLPNNVEVSVDYTAWVMPSPKSPFPCDLNWL
DNQDDGSLGPIIKDLSSTLACVLAKRWNVGDKQINKGLHYIESNFASVT
DKNOASPFQDIIFPQMLBYAKDLIKLPLNOLHLSVMLHERELELRCHSNGREAYL
AVISGLGNLNDNMWVKYQMGKSLFNPSATASVLJHQNAGCLHILTSLLDFGN
AVPTYPIDLYRLSNVDLERLGRHFWEIQNVLDIETRCWVGQVQIFMDVVTTC
ALAFVRLNSGEVSDPLAKITKGDYNNSPKPFKDVYTSLEVYKASQIITYBELA
FREQNLTSLPSNLSNYILKEVDLALFPFNGSLERMSTRNIEHYNLNHLRIKLT
TYSNNISNKDVLKIAVQDPNCSQIYCEELKDLRWVVENRDLKLFAROKTAYCF
SAASFLSPDLKARISWAKSLITTVIDDFDVGSDMDLVNFVHIIIEKNVYNEND
CCSEEVGLVFLAKDAVCGDKAPKIQERNITSHVIEIWLVLKSMLEALWAKDGS
IPTINEYMGVPSFALGIVLPITFLYFLKLSSEVVQSEYHKLYEVNSTQGRIMND
IHSFKREKAGKLNALVSDGKSGVEEVBEEMKILTKSQRKEMKMLVLETKGSV
VPRVCKDVFNNVNCVNLNLFYATDGDGTGNAILDVVKBIITYEPVSHELI"

Query Match 51.2%; Score 1428.6; DB 8; Length 2594;
Best local Similarity 75.8%; Pred. No. 9.5e-270;
Matches 1833; Conservative 0; Mismatches 529; Indels 57; Gaps 3;

ORIGIN
```

1255 AGTTAGCTTTAGAGACAAACTTGA-----CAAGTTATCTCC 1293
1354 CTGATTCGAACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAAAATGCTCTTAAGTTCC 1413
1294 CTTCAAGTAAACAGCTCTCTAAATATATATCTCAAGAGAGGTGGATGATGCTCTTAAGTTCC 1353
1414 CTATCAATACGGTTTGAAGCGCAATAACACTAGACGAAATATACAGCTTTACAAATGTAG 1473
1354 CTTTAAATGGCAGCTAGAACGCAATGCCACTAGAGAAACATAGAGCAATACAAATTAA 1413
1474 ACAATAACAAGAAATCTGAAAACCTACATATCACTCATCAAAATATTAGTAACAATGATACC 1533
1414 ACCATACAAGAAATCTTGAAGAACTACATATAGCTCATCAACATATTAGCAACAAGATTACC 1473
1534 TAAGTTGGCTGTTGAAGATTTCTACACCTGCCAATCTATTATTCGTTGAAGAAATAAAG 1593
1474 TAAAGTTAGCTGTGCAAGATTTCAATGAATGCCAATCTATCTATTGTTGAAGAACTAAAG 1533
1594 GTCTTGAAGAGTGGTGTGAGAAATAAGTTGGACCACTCAAGTTTGTAGGCAAAAGA 1653
1534 ATCTTGAAGAGTGGTGTGAGAAACAGATTAGCAAGCTCAAAATTTGCTAGACAGAGA 1593
1654 CCGCTACTGTTAATTTCTGTTGCTGCAACACTTTCTGCTCCCGAATTTATCAGATGCC 1713
1594 CAGCTACTGTTAATTTCTGCTGCTTCAATTTCTTCTTCTCCCTGATTTATCAGATGCC 1653
1714 GTATTTCAATGGCCAAAAGGCGATATACTACAGTAGTGTGATGACTTTTTTGAATACG 1773
1654 GAATTTCAATGGCCAAAAGTAGCATCTCTCACTACAGTAATTTGATGACTTTTTTGAATG 1713
1774 GTGTACAATCGATGAATTTGACCAACCTGATTTCAATGTTTGAAGAAATGGAATCTAGATG 1833
1714 GAGGATCTATGGATGACTGTGTAACCTTTGTTCACTCATTTGAAAATGGAATGTAACG 1773
1834 TCGAAGGATGTTGTTTCAGAGCAATGTTCCGATTTTATTTTATAGCAATTAAGAATGCAA 1893
1774 TTGAAAATGATGTTGTTTCAGAGCAATGTTGCGTTTATTTTATAGCACTTAAAGGATGCG 1833
1894 TCTGTTGATGGAGATGAGCTTTTAAATGCGAGCGCGGATGTAAGTAACTAGCATGTTA 1953
1834 TCTGTTGATGGAGACAAAGCTTTTAAAGATTCAGAAACGCAATATACAAAGCATGTTA 1893
1954 TTCAAACTTGTTTGGAACTAATGATAGTATGTTGAGAGAGCTATATGACACAGAGATG 2013
1894 TTGAAATTTGTTTGAATTTGTTGAAGAGTATGTTGAGAGAGCTATATGCGGCGAAGATG 1953
2014 CTTATGTGCCAACATTAATGAATATATGGAAGACGCTTAGTGTCAATTCATTTAGGCC 2073
1954 GTTCAATTTCCAACAATAAATGAATATATGGAAGATGTTAGTGTCTCAATTTGCTTTAGGCC 2013
2074 CGATTGTCAAGCGCGCTATTTACTTTTGTGGGCGCCAAATATATCAGAGAGATTTTGAAA 2133
2014 CGATTGCTCCCTACTTTTACTTTTATAGGGGTAAATTTGTCGAGGAAGTTGTTGAGA 2073
2134 GCTCTGAATATCATAACTATTTAAGCTAATAGCAGCAGCGGTCGATCTTCAACAGATA 2193
2074 GCTCCGAGTATCAAGCTTTATGAGGTTATGAGCACTCAGCGTCTGCTCATGAATGATA 2133
2194 TCCATAGCTTCAAGAGGAATTTAAGAGAGCAATTAACCGCGGTAGCATTGATTTGA 2253
2134 TCCATAGCTTCAAGAGGAAAGAGCGCGGGAATTTGAATGCTGTGCGCATTTGATACATGA 2193
2254 GTAAACGGAAGAGTGGGAAAGTGGAAAGAGGTTTGTGAGAGAGATGATGATGATTA 2313
2194 GTGATGGAAGAGTGGGAGTGTGGAAGAGAGGTTTGTAGAGAGAGATGAAGATTTTGACA 2253
2314 AAAACAAGAGAGAAATTAATGAATTTAATTTTGAAGAAAATGTTAGCATTTGTTCTTA 2373
2254 AAAAGTCAAGAGAGAGATGATGAAATTTAGTTTGGAAAACAAAGGAAGTGTGTTTCCAA 2313
2374 GAGCTTGTAAAGATGCATTTTGGACATGTTGTCAGCTGTGAAATTTTATTTTACCAACG 2433

2314 GAGTATGCAAGAGATGATTTTGGAAACATGTGCAATGTGTTGAATCTTTTTTATGCGACAG 2373
2434 ATGACGGTWTACTTGGAAACACGATTTCTTGATATCTGTGAAGAGCATCATTTTACAACCGT 2493
2374 ATGATGGTTCACCTTGGAAATGCCATTTCTGATGTTGCAAGGAATTTATATATGACCTG 2433
2494 TGTGCTTGTGCAATGAAATGAGAACAAAGGTAATTTGGCCTTTTAAAGATGATAAATAAT 2553
2434 TGTCCCATGAGTTGATATAATCATTTGTTATACCAATTCGTATTATATTATTAGCATCTC 2493
2554 ATCATGCTTTAGCGGGGTACTTGTGTAGTTGTATATAAAGGTTGATGTTGATATAATA 2613
2494 AAAAATGAAACACAGCATACTAGTATTCTAGATCTACTCATGTTTATTATTATTATTTT 2553
2614 AAGGTAATAGTAAATCAAT 2632
2554 AATATGTTTGTAGCTAT 2572

AB045310 2450 bp mRNA linear PLN 06-JUN-2002
Cucumis sativus CsKS1 mRNA for ent-kaurene synthase, complete cds.
AB045310.2 GI:21326756
Cucumis sativus (cucumber)
Cucumis sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
1
Shirai, K., Fujino, K. and Masuda, K.
Cloning and sequencing of an ent-kaurene synthase cDNA from
cucumber
Unpublished
2 (bases 1 to 2450)
Shirai, K., Fujino, K. and Masuda, K.
Direct Submission
Submitted (28-JUN-2000) Kikuko Shirai, Graduate School of
Agriculture, Hokkaido University, Laboratory of Horticulture
Science, Kitaku Kitagunishi, Sapporo, Hokkaido 060-8589, Japan
(E-mail:kik@res.agr.hokudai.ac.jp, Tel:81-11-706-3875(ex.3875),
Fax:81-11-706-4937)
On Jun 5, 2002 this sequence version replaced gi:11862932.
COMMENT
FEATURES
Location/Qualifiers
1..2450
/organism="Cucumis sativus"
/mol_type="mRNA"
/db_xref="taxon:3659"
/clone="CsKS1"
1..2450
/gene="CsKS1"
23..2380
/gene="CsKS1"
/codon_start=1
/product="ent-kaurene synthase"
/protein_id="BAB19275.1"
/db_xref="GI:11862933"
/translation="MNLSPRLNLCPTASSASLFPGLDVGTKTKTGCALPEETKERI
KKLFWNELSISAYDTAWAVNPFNSLNKPLFPPECINWLDHNPQSGWGLHDHQL
VMKATLSTLACVLTLEKRWIDGDDHMSKALSFTKSNISATDENQSPFVGDIIFPGM
IETAKDLNLNPLASNNVDALVQKLELRSCCSNREGKAYLAYVSEGGTGLQDWEM
VMRYORNGSLFSPSTTAVAFWHRNDGCFNYLRSVLQKPFHSPVPAIYPLDIYARH
MVDLQKLGIDGHFKDRIRSLDRTYSCWNGENIFLDASTCMAFMRLRVEGYDVS
SDQLTORSEGLFNSCLGGHLKDFASLEFKASOLIIYPDEFLELNLSRFLNHG
LSSGSVSRTERLIVKQEAIVNAFEFPNLSRLERLSNRKALRESYSGDIVRSKTAACL
NFGHQDFLELAEVDFTLQGHRRKLEKLEKLVLENKLDKLFKARQKLACIYFSAAT
LITPELCHDLRLSWAKGLTTFVDDFDVGGSEELVNLILQLEKWDASGTYCSKE
VEIFLALHSTICIGIKKALFWQGRSVNRNVIDTLALLESMRKEAWLKNKVPVSLD
EYEMGVVVALGPVLPVTFVFGKLPPEEIVGNCYKLFKLMSTSGRLINDRTFD
RESSEGLNLSLYMISAGGLTKEATEAMKGDVDRTRRELLRLVLQENSTIPRACK
DLFWKMSCVVHLFRKDDGTSHELMSAKALFEQPMVLDLNNK"

| ORIGIN | | | |
|--|------|---|------|
| Query Match 29.0%; Score 809.4; DB 8; Length 2450; | | | |
| Best Local Similarity 62.0%; Pred. No. 2.1e-148; | | | |
| Matches 1410; Conservative 0; Mismatches 811; Indels 54; Gaps 6; | | | |
| QY | 267 | GGTGGACAACTAATCCACATCTGATCTGATCAATGATACCAACCAAGAACGGATCCAAAA | 326 |
| DB | 101 | GGCAGAAAACTAAACTGAGCTTTGGCTTTGAAGAAAACAAAGAAAAGAAATAAAAAA | 160 |
| QY | 327 | CAGTTTAAAAATGAGAAATTTCTGTTCTTCAATGACACAGCATGGGTAGCCATGGTC | 386 |
| DB | 161 | TTGTTCAAAAATGTTGAATCTTCAATTTCTGATATGATCTGATGGTGGCAATGGTC | 220 |
| QY | 387 | CTTCTCCAAACTCACCAAAATGCGCTTGTTCCTGAGTGTCTCAATGGTTTAATTAAT | 446 |
| DB | 221 | CTTCTCCAAACTCTTAAATAAACCTCTTTTCTGAGTGCATAAACTGGGTATTAGAT | 280 |
| QY | 447 | AATCGACTTAATGATGGTTCATGGGTCTTGTAAATCACACTCAATAATCAATCAACCG | 506 |
| DB | 281 | CATCAAAACCCCTGATGGTCTATGGG-----CATCTCCATGACCATCAGTTG | 328 |
| QY | 507 | TGCTTTAAAGATTCCTATCTTCAACATTAGCATGTAATGTTGCAATTAATAAGATGGAAT | 566 |
| DB | 329 | GTGATGAAGCCACTCTCTTATCCATTTAGCAATGATCTTACTCTTAAAGCATGGAT | 388 |
| QY | 567 | GTGCGGAGATCAAAATAAAGGTCTAAGTTTATGAGTCAAACTTGTCTTCAGCT | 626 |
| DB | 389 | ATCGGTGATGATCATATGAGCAAGCCCTTAGTTTATCAAGTCTAATATAGCTTTCAGCT | 448 |
| QY | 627 | ACTGAAAAAAGTCAACCATCTCCATTTGTTTGGATCAATATTTCTGTTTCTGATG | 686 |
| DB | 449 | ACTGATGAGATCAAGCTTCTCTGTTGGGATTTGACATAATTTTCCCTGGTATGATGAG | 508 |
| QY | 687 | TATGCGAAAACTTGACATAAAGCTCTTTCAAAAACAAACAGATTTAGTTTGTATGCTA | 746 |
| DB | 509 | TATGCTTAAAGACTTGAATTTGAACTCTACCTTGGCATCAATGAACTGGATGCTTTGGTT | 568 |
| QY | 747 | CATAAGAGGAATTTGAGCAAAAAAGAT-----GCCATTCAATGAGATGATGATGATAC | 800 |
| DB | 569 | CAAAAGAAAGTTTGGAGCTTAGAAGCTGCTGTAGCAACTCTGAAGGAGGAAAAAGCCTAT | 628 |
| QY | 801 | TTGGCGTATATCTCTGAAGGATCTGGTAAATTTATATGATTTGGAATGATGGTGAAGAAAT | 860 |
| DB | 629 | TTAGCGTATGTTTCAAGGAATTTGAAAAATTAACGACCTGGGAAATGGTCAAGCATAT | 688 |
| QY | 861 | CAGATCAAAAATGCTGCTGTTTCAACTCAACATCAGCAACAGCTGCTGCTTTCAATTAAT | 920 |
| DB | 689 | CAAGGAAGAACGGATCACTGTTTATGTTTCTTCCATCCACACGCGAGTGGCTTTTATGCAC | 748 |
| QY | 921 | CATCAAAAATCCTGTTGCTTAAATTTAAATTTCACTTTTGGACAAAGTTTGGTAATGCA | 980 |
| DB | 749 | AGAAATGATGATGCTGTTTAAATTAACCTTCGCTCAGTCTTACAAAAGTTTTCATAGTTCA | 808 |
| QY | 981 | GTCCCAACAGTTTATCCTCATGATTTATTTATCGACTTCTATGTTTGGACACAAATTTGAA | 1040 |
| DB | 809 | GTTCGCCCAATATATCCTCTGATATATATGCTGTTTACACATGGTTTGAAGCTTTCAA | 868 |
| QY | 1041 | AGATTAGAAATTTACACACCAATTTTCAAGATGGAAATTAATAATGTTTATGATGAAACATAC | 1100 |
| DB | 869 | AAACTGGGGATGATGAGGCCATTTCAAGATGAGATTAAGATGATTAATGATGAACATAC | 928 |
| QY | 1101 | AGATGTTGGGTGGAACGAGATGAGCAAAATTTCAATGATGTTTGAACATGCTTTTAGCC | 1160 |
| DB | 929 | AGCTGTTGGATGCAAGGGGAGGAAAAATATTCCTAGATGCTTCAACTTGTGCAATGGCC | 988 |
| QY | 1161 | TTTCGGTTTAAAGGATCAATGGGTGATGAAGTTTCCCGAGATCCATTTGGCTGAAATTAAT | 1220 |
| DB | 989 | TTCCGGATGTTACGTTGTTGAAGGATATGATGTTTCTTCAGATCAATTAATGACTCAATCTCA | 1048 |
| QY | 1221 | AA-----TGAATTAGCTTTGAAAGACGAATATGACAGCTCTTTGAA | 1259 |
| DB | 1049 | GAAGGTCTCTTTTCAAAATTTGCTCGAGGACATTTTAAAGAACTTTAGTGCCTCTACTAGAG | 1108 |

| | | | |
|----|------|---|------|
| QY | 1260 | ACATATCATCGGTCCACA---TATATTATACCAAGAGGATTTATCTTTGGAACAAATC | 1316 |
| DB | 1109 | TTATTTAAGGCCCTCCAGATTAATCATTTATCCGATGAGTTTATCTTGGAAAAATATAAC | 1168 |
| QY | 1317 | TTGAAGTCAGCTGATTTCTCCTCAAGAGATTAATATCCACTG-----ATTCAACAGG | 1367 |
| DB | 1169 | TCCTTGGACTAGTCTGTTTCTCTGAATCATGGAATTAATCTAGTGGTTCAGTTTCTTGATAGA | 1228 |
| QY | 1368 | CTTTCTTAATTAATTCACAAAGAGGTGAAAAATGCTCTTAAGTTTCCCTATCAATACCGGT | 1427 |
| DB | 1229 | ACTGAGAGACTCGTGAACAAAGAGCGAGTTAATGCTTTTCGAGTTCCCTATAATTTCAACT | 1288 |
| QY | 1428 | TTAGAACGCTATAAACACTAGACGAAAAATATACAGCTTTCAATGATGACAAATCAAGAAAT | 1487 |
| DB | 1289 | CTGGAACGCTTATCAATAAGCGAGCACTGAAAAATTAACAGTGAGACATTTGTGAGGAT | 1348 |
| QY | 1488 | CTGAAAACTACATATCACTCATCAAAATATATAGTAACACTGATTAACCTAAAGTTCGCTGT | 1547 |
| DB | 1349 | TCAAAAAACAGCATATGCTGCTTAAATTTTGGTTCATCAAGATTTCTTGGAACTTCTCTGTA | 1408 |
| QY | 1548 | GAGATTTCTACACTGCTCAATCTTATTTATCGTGAAGAAATTAAGAGGTCTTTGAAAGTGG | 1607 |
| DB | 1409 | GAGATTTTCAATACCTCTCAAGGCATACATCGAAGGAACTGAAAGAGCTTTGAAAAATGG | 1468 |
| QY | 1608 | GTGGTAGAGAAATAAGTTTGGACCACTCAAGTTTGTCTAGGCAAAAGACCGCTTACTGTTAT | 1667 |
| DB | 1469 | GTCTCGAAAAACMAATTTGCAAGTTTGAATTTTGGAGACAGAGTTAGCGTACTGCTAT | 1528 |
| QY | 1668 | TTCTCTGTTGTGCAACACTTTTCTCCGAAATTAATCAGATCGCGGATTTTTCATGCGGC | 1727 |
| DB | 1529 | TTTTCTGCTGACGCAACCTTAACTCTCCAGAACTTTGTGATGCGCGCTTATCATGCGGCA | 1588 |
| QY | 1728 | AAAAATGCAATTAATTAACACTAGTGTGATGACTTTTGTGATATCGTGGTACAATTCGAT | 1787 |
| DB | 1589 | AAAAATGCGGTACTCACAAACCGTGGTTGATGATTTCTTTGATTTCTTTGAGGATCTGAAGAG | 1648 |
| QY | 1788 | GAAATGACCAACCTGATCAATGCTGTGTAATAATGGAATGTAGATGTGCAAGGATTTGT | 1847 |
| DB | 1649 | GAAATGTTAACTTATACAAATTTGTTGGAATAAGTGGATGCCAGTGGGAAAAACGGGTAC | 1708 |
| QY | 1848 | TCCTCAGAGCATGTTCCGATTTTATTTTAGCAATTAAGATGCAATCTGTTGGATTTGGA | 1907 |
| DB | 1709 | TGTTCCAGAGGTTGAGATTTATTTCTTTCACATTCATAGACAAATTTTGTGAATATGGA | 1768 |
| QY | 1908 | GATGAGCTTTTAAATGCGCAACGCGCATGTAACCTAGCCATGTTTATCAAACTGCTGTTG | 1967 |
| DB | 1769 | AAAAAAGCTTTTACCTTGGCAAGACGCGCTGATGAGGAATTTATCGATTTTGTGTTG | 1828 |
| QY | 1968 | GAACTAATCAATAGTATGTTGAGAGAAAGCTATATGGAAGAGATGCTTTATGTGCCAACA | 2027 |
| DB | 1829 | GCTTTGCTCGAGTCAATGAGGAAGAACTGAATGTTTGAATAAAGGTAGTGCCATCA | 1888 |
| QY | 2028 | TTAAATGAATATATGAAAAACGCTTACGTGTCATTTGCAATTAGCCCGGATTTGCAAGCG | 2087 |
| DB | 1889 | TTGGATGAGTACATGAAAAATGCTATGATCAATTTGCTTTGGGACCTATAGTCTTCCA | 1948 |
| QY | 2088 | GCTATTTTCTTTGCGGGCCAAATTTATCAGAGGAGATTTGTCAAAAGCTCTGAATATCAT | 2147 |
| DB | 1949 | ACGCTCTACTTTTGGACCTAAGCTTCCAGAGGAATTTGTTGGAATTTGTGAATCAAG | 2008 |
| QY | 2148 | AATCTATTTAAGCTAATGAGACGCGGGTTCGACTTCTTAAACGATATCCATAGCTTCAAG | 2207 |
| DB | 2009 | AAGCTCTTTAAGCTGATGAGCACTTCTGGCGGCTTCTGAATGATACTCGAACTTTTGAT | 2068 |
| QY | 2208 | AGGGAATTTAAGGAAGCAAAATTAACGCGGTGAGCATTTGCAATTTGATTAACGGAGAACT | 2267 |
| DB | 2069 | AGAGAGTCCAGCGGAAGGAAATTAATGCTCTCTATACATGATTTAGTTCGCGGTGGT | 2128 |
| QY | 2268 | GGGAAGTGAAGAGAGGTTTGGAGGAGATGATGATGATGATTTAAATAAACAAGAGGAAA | 2327 |
| DB | 2129 | AAGCTCACCAAGAGGAGGCCACTGAAGCAATGAAGAGATGTTTGTAGACGAGAGA | 2188 |

| QY | 2328 | GAATTAATGAAATTAATTTTGAAGAAATCGTAGCATTTGTTCTAGAGCTTTGTAAGAT | 2387 |
|-----------------------|---|--|--------------------|
| Db | 2189 | GAACCTGTTGAGATTAGTTTTCAGGAGAACAGTACAA---TTCCAAGAGCTTTGTAAGCAT | 2245 |
| QY | 2388 | GCATTTTGGACATGTGTCACGTGTTGAATTTTTCACGCAACGATGCGGTTTACT | 2447 |
| Db | 2246 | TTGTTCTGGAATAAGAGCTGTGTTGTCATCTATTTTACAGGAAGATGATGGGTTTACA | 2305 |
| QY | 2448 | GGAAACACGATCTTTGTATPACTGTGAAGGACATCATTTACACCCGTTGTTGCTTG | 2502 |
| Db | 2306 | TCCCATGAGTTGATGAACCTCTGGGAAGCCTTATTGACACACCATGGTCTCG | 2360 |
| RESULT 5 | | | |
| LOCUS | E12936 | 2658 bp | DNA |
| DEFINITION | | cDNA encoding ent-kaurene synthase A. | |
| ACCESSION | E12936 | | |
| VERSION | E12936.1 | GI:3251767 | |
| KEYWORDS | JP 1997104698-A/1. | | |
| SOURCE | unidentified | | |
| ORGANISM | unclassified. | | |
| REFERENCE | 1 (bases 1 to 2658) | | |
| AUTHORS | Kamiya, Y. and Yamaguchi, S. | | |
| TITLE | KAURENE SYNTHETASE | | |
| JOURNAL | Patent: JP 1997104698-A 1 22-APR-1997; | | |
| COMMENT | RIKAGAKU KENKYUSHO OS Cucurbita maxima L. PN JP 1997104698-A/1 PD 22-APR-1997 PF 09-OCT-1995 JP 1995261147 PI KAMIYA YUJI, YAMAGUCHI SHINJIRO PC C07K14/415, C07H21/04, C12N1/21, C12N9/88, C12N15/09, C12N1/21, PC C12R1.19), PC (C12N9/88, C12R1.19); CC CC CC CC CC CC FH FH FH FH FT FT FT FT | | |
| FEATURES | | | |
| source | 1. .2658 138. .2507 | /organism='Cucurbita maxima L.' FT CDS /product='ent-kaurene synthase A'. | |
| ORIGIN | | | |
| Query Match | 28.0%; | Score 782.2; DB 6; | Length 2658; |
| Best Local Similarity | 61.3%; | Pred. No. 4.4e-143; | |
| Matches 1390; | Conservative 0; | Mismatches 823; | Indels 54; Gaps 6; |
| QY | 275 | AACTAATCCCACTATCTGATCTGATCATTTGATACAAACGAGATCCAAAAACAGTTTAA | 334 |
| Db | 236 | AACTAAACTGGAGCTTTTGCACTTTTGAAGAGAACAAAGAGAAATTAAGAAATTTTCGA | 295 |
| QY | 335 | AAATGTAGAAATTTCTGTTTCTTCATATGACACAGCATGGGTAGCCATGGTCCCTTCTCC | 394 |
| Db | 296 | CAAGGTTGAACCTTTCAGTTTCTGCATATGATCTGATGGGTGGCAATGGTTCCTTCTCC | 355 |
| QY | 395 | AAACTCACCAGATCGCTTGTTTCCCTGAGTGTTCAATTTGGTTTAATTAATCAATCAGCT | 454 |
| Db | 356 | AAACTCTCAACCAACCTCTTTTCCCGAGTGATTAACCTGGGTATTAGATATCAACA | 415 |
| QY | 455 | TAATGATGTTTCATGGGTCGTGTTAATCAACATCAATATCAATCAATCAATCAATCAATCAAT | 514 |
| Db | 416 | TGCTGATGCTCATGGGCCCTACTCCAAAC-----GATCAGTTGCTGATCAA | 463 |
| QY | 515 | AGATTCTCTATCTCAACATGAGCATGTTTGTGCATTAAAGAGATGGAATTTGGGA | 574 |

| | | | |
|------|----|---|------|
| 464 | Db | GGCCAAATCTCTTATCTACATTAGCATGTGTTCTTACTCTTTAAAGCGGTGGAATATTGGGCA | 523 |
| 575 | QY | AGATCAAAATAAAAGGTCCTAAGTTTTTATTTGAGTCAAAATCTTGCTTCAGCTACTGAATA | 634 |
| 524 | Db | TTTATTTATTTGAGCGAGGCCCTTGATTTTATCAAGTCTAATATATAGCTTCAGCTACTGATGA | 583 |
| 635 | QY | AAAGTCAACCATCTCCCAATTCGGTTTTGACATCATATTTCTCTGGTTTTCGTTTGAATATCGAA | 694 |
| 584 | Db | GAACCAAGGTTCTCCGGTGGGATTTGACATTAATTTCCCTGGCATGATTTGAGTATGCTTAA | 643 |
| 695 | QY | AAACTTGGACATAAAAGCTCTCTTCAAAAACAAACAGATTTTATGTTTGAATGCTACATAAGAG | 754 |
| 644 | Db | AGACTTTGAATTTGAATCTACCTTGGCACCGACGAACTGGATGCCCTTGTTTCGAAAGAA | 703 |
| 755 | QY | GGAAATTTGGAGCAAAAA-----AGATGCCATTCAAATGAGATGGAATGATCTTTGGCGTAA | 808 |
| 704 | Db | AGAGTTTGGAGCTGAGAAAGCTGCAAGAAGCACTCTGGAAGTGGAAAAGCCCTATTATAGCGTA | 763 |
| 809 | QY | TATCTCTGAAGGACTCGGTAATTTATATGATTTGGAATATGGTCAAGAAATATCAGATGAA | 868 |
| 764 | Db | TGTTTTCAGAAGGAATTGGAAAGTTACAGACTGGGNATAGGTATGCAATATCAAGAAGAA | 823 |
| 869 | QY | AAATGGTCTGTTTTCAACTCAACATCAGACAAAGCTGCTGCTTTCAATTAATCATCAAAA | 928 |
| 824 | Db | GAATGGATCACTGTTTAATTTCTCCATCCACTACGGCAGCGGCTTTTATGCAATAGAAAATGA | 883 |
| 929 | QY | TCTCGTGTCTTTAAATTAATTTAAATTTCACTTTTGGACAAGTTTGGTAATGCACTCCCAAC | 988 |
| 884 | Db | TGATGGCTGTTTGAATATCTTCGGCTCACTCTTACAAAAGTTTGAAGTCTCAGTTCCTCAC | 943 |
| 989 | QY | AGTTTATCCTCATGATTTATTTATCCGACTTTCTATGTTTGACACAAATGAAAGATTAGG | 1044 |
| 944 | Db | AATAATCTCTTGATATATATGCTCGATACACATGGTTGATAGCTTTCAAAAATTCGG | 1003 |
| 1049 | QY | AAATTTACACCATTTTCAGGTGGAATAATAAAGTTTTTACATGAAACATACAGATGTTG | 1108 |
| 1004 | Db | AAATGCTCGGCATTTCAAAGAGGAGATTAGAAGCGTATTAGATGAAACTTACAGGTGTTG | 1063 |
| 1109 | QY | GGTGGAAACGAGATGAGCAAAATATCATCGATGTTGTAAATGCTGTTTAGGCTTTTCGGTT | 1168 |
| 1064 | Db | GATGCAGAGAGAGGAAATATATCTCTAGATGCTTCAACTTTGTGCAATGGGCTTTCCAAAT | 1123 |
| 1169 | QY | ATTAGAGTCAATGGGTATGAAGTTTCCCCAGATCCATTGGCTGAAATTTACTAA----- | 1222 |
| 1124 | Db | GTATAGCTGTTCGAAGATATGATGTTCTTTCAGACGAGTTGACTCAATTTTCAGAAATAT | 1183 |
| 1223 | QY | -----TGAAATAGCTTTGAAAGACGAATATGACGCTTTGAAACATATCA | 1267 |
| 1184 | Db | CTTTTCCAAATTCGCTTGAGGATATTTAAAGACTTCGGTGCCTCGCTGGAGTTATATAA | 1243 |
| 1268 | QY | TGCGTCAATATATATACCAAGAGGATTTATCTTCT---GGAAAAACAATCTTTGAAGTC | 1324 |
| 1244 | Db | GGCTCTCAGATTAATCAGCACCCCGGATGAATCTGTTCTGAAAAATATAAATCTTTGGAC | 1303 |
| 1325 | QY | AGCTGATTTCTCAAAGAGATAATATCCACTGATTTCAAACAGGCTTTCTAAA----- | 1376 |
| 1304 | Db | TAGTCGTTTCTCTGAAGCATGGAATTAATGATGATTTAGTTTGGTCTGATAGAACCGATAG | 1363 |
| 1377 | QY | TTAAATTCACAAAGAGGTGGAATAATGCTCTTAAAGTTCCTTATCAATACCGGTTTAGAACG | 1435 |
| 1364 | Db | TGTTGTTTAAACAAGAGGCTGTAAATGCTCTTGAGTTCCCTCTAATGCAACTCTAGAAGC | 1423 |
| 1436 | QY | CATAAACACTAGACGAAATATACAGCTTTTACAATGTAGACAATACAGAAATTCGAAAC | 1495 |
| 1424 | Db | CCATTAAGTTAAGGGCAATGGAAGTTTACAGTGGAGACATTTGTGAGGATTTTCAAAATC | 1483 |
| 1496 | QY | TACATATCACTCATCAATATTAAGTAAACACTGAATTAACCTTAAGGTTGGCTGTTTCAAGATTT | 1555 |
| 1484 | Db | GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGGAATTTGCTGTAGAGATTTT | 1543 |
| 1556 | QY | CTACACCTGCCAATCTATTTATCGTGAGAAATTAAGGTTTGAAGGTTGGTGGTAGA | 1615 |
| 1544 | Db | CAATACCTCCACGCAATTCATCTTAAAGAACTGGAGAGCTTCAAGAATGGGTGGTTGA | 1603 |

| | | | | |
|------------|--|---|------|---|
| QY | 1616 | GAATAAGTTGGACAGCTCAAGTTTGCTAGGCAAAAGACGCCCTACTGTTATTTCTCTGT | 1675 | rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita. |
| Db | 1604 | AAACAATTTGACGAGTTGAATTTTTCAGACTGCACCTAGGCTACTGCTATTTTGTCTGC | 1663 | 1 (bases 1 to 2658) |
| QY | 1676 | TGCTGCAACACTTTCGCTCCCGAATTTATCAGATGCGCGTATTTTCATGCGCCCAAAATGG | 1735 | Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N. and |
| Db | 1664 | GCCAGCGACCTTACTCATCTGAACTTCATGCTCGCATGATGCGGCACAAAATGG | 1723 | Kamiya, Y. |
| QY | 1736 | CATATTACACAGTGTGATGATCTTTTGTATCGGTGGTCAATCGATGAATTTGAC | 1795 | Molecular cloning and characterization of a cDNA encoding the |
| Db | 1724 | TGTGCTCAGCAGCGGTGTTGATTTTCTATGATGGTGGAGTCTGAAGAGAAATGGA | 1783 | gibberellin biosynthetic enzyme ent-kaurene synthase B from pumpkin |
| QY | 1796 | CAACTGATTCATATTTTGAATAATGAATGTAGATGTGCAAGAGATTTGTTTGA | 1855 | (Cucurbita maxima L.) |
| Db | 1784 | TAACTTTATAGAAATTTGTTGAAAGTGGATCCTGATGGGGAAGTGGTACTGTTCCAA | 1843 | Plant J. 10 (2), 203-213 (1996) |
| QY | 1856 | GCATGTTCCGATTTTATTTTGAATTAAGATGCAATCTGTTGGATTTGGATGAAGC | 1915 | 96367664 |
| Db | 1844 | GGACGTTGAGATTGATTTCTTGCACTGCACAGCACAGTTTGTGAATAGGAAGAGC | 1903 | 8771778 |
| QY | 1916 | TTTTAAATGCAAGCGCGCATGTAACCTAGCATGTTATTCAACTTGGTTGGAACTAAT | 1975 | 2 (bases 1 to 2658) |
| Db | 1904 | TTTAGTATGGAAGACGACGATTTATGAGGAATGTTATCGATGCTGTTGGCTCTGCT | 1963 | Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N. and |
| QY | 1976 | GAATAGTATGTTGAGAGAGCTATATGCAAGAGATGCTTATGTCGCAACATTAATGA | 2035 | Kamiya, Y. |
| Db | 1964 | GAACTGTATGAGAAAGGAGCTGAATGTCGCAAAATAGGTAGTGCCTCAATCGGTGA | 2023 | Direct Submission |
| QY | 2036 | ATATATGGAAGACGTTTACGTTGCTATTTGCAATAGCCGATTTGCAAGCCGGCTATT | 2095 | Submitted (25-DEC-1995) Shinjiro Yamaguchi, The Institute of |
| Db | 2024 | ATATATGGAAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 2083 | Physical and Chemical Research, Hiroswa2-1, Wako-shi, Saitama |
| QY | 2096 | CTTTGTTGGGCCCCAATATATCAGAGAGATTTGTTGAAAGCTCTGAAATCATTAAT | 2155 | 351-01, Japan |
| Db | 2084 | CTTTGTTGGACCTTAACTCTCAGAGAAATGATTTGGAAGCTGTAATACCAAGATT | 2143 | Location/Qualifiers |
| QY | 2156 | TAACTGTAATGAGACGCGAGGTCGACTTCTAAACGATATCCATAGCTTCAAGAGGA | 2215 | 1. .2658 |
| Db | 2144 | TAACTGTAATGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 2203 | /organism="Cucurbita maxima" |
| QY | 2216 | TAAAGAGGCAATTAACGCGGTAGCATTCATTTGATTAAGGAGAGAGAGGAGGAA | 2275 | /mol_type="mRNA" |
| Db | 2204 | CAAAGAGGAAAGCTGAATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 2263 | /db_xref="taxon:3661" |
| QY | 2276 | GGAGAGAGGTTGTGAGAGATGATGATGATTAATAAACAAGAGGAAAGAAATTAAT | 2335 | /tissue_type="immature seeds" |
| Db | 2264 | CAAAGAGGAGGCAATTAAGCAATTAAGGAGATTTGAGAGGCGATTAAGAGAGCT | 2323 | 138. .2507 |
| QY | 2336 | GAAATTAATTTTGAAGAAATGTTAGCATTTCTTACGATTTGTAAGATGATTTGTT | 2395 | /function="ent-kaurene synthase B in gibberellin |
| Db | 2324 | GGGGTTAGTTTTCAGGAGAACTACAA--TTCCAAGAGCTTTGTAAGGATTTGTTCTG | 2380 | biosynthesis" |
| QY | 2396 | GAACTGTGTCAGTGTGATTTTTCAGCAAGATGAGAGAGGTTTACTGGAACAC | 2455 | /note="terpene cyclase" |
| Db | 2381 | GAAATGATGCTGATTTGATTTTATTTACAGAGAGATGAGGTTACATTTCAATAG | 2440 | /codon_start=1 |
| QY | 2456 | GATTTCTGATGTTGAGGAGACATCATTTACAAACCCGTTGTTGTTG 2502 | | /product="ent-kaurene synthase B" |
| Db | 2441 | GTTGATGAACACTGTAAAGCCATGTTTGAACACCCATGATCTGG 2487 | | /protein_id="AB39482.1" |
| RESULT 6 | | | | /translation="MYSRPTGVARFAAASSSSSASLPPGVDDVTTTKTGALHPEET |
| LOCUS | CMU43904 | | | /db_xref="GI:1431870" |
| DEFINITION | Cucurbita maxima ent-kaurene synthase B mRNA, complete cds. | | | /translating="MYSRPTGVARFAAASSSSSASLPPGVDDVTTTKTGALHPEET |
| ACCESSION | U43904 | | | KERIKLFDKVELSVADTAWAVPSPNSINQPLFECINWILDSQADSGWGLH |
| VERSION | U43904.1 | | | NDQLLKANLSTLACVLTAKWNTGHDHMSKALDFIKSNTASADENQSRVGFII |
| KEYWORDS | | | | PFMEYAKDLNLNPLAPTVNDALVRKLELRCSRNSGCKAYLAVVSGIGKIQ |
| SOURCE | Cucurbita maxima (winter squash) | | | DWDVMVQYKNGSLFNSPSTAAAFMRNDDGCFDYLRSLLQKFDGSPVPTVLDIY |
| ORGANISM | Cucurbita maxima | | | YKHSLSQSVMSDRDTSVVKQAVNALFPYNATLRLISKRAMSESGDVIIRISKP |
| | Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; | | | LKHNVDLQKFGIARHFKERISVLDITRCWQOGENIFLDASTCAMAFLMVEG |
| | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; | | | YACINPHQDFLEAVDFNTLQTHLELELQWVVENKLELFFELHGYCYFA |
| | | | | AAATDPELHDARITAWQGVLTVDVDDFYDGGSEELDNLIELVEKWDPEVGY |
| | | | | CSKDEIVFLAHSTVEIGREALVWQGSVMRNVLDGWLALLKVMKEAENSTKNV |
| | | | | PMGSMQEAHVSPALGPIILPMLFPVPGKSEEMIGSCYOKLYKLMSTAGRLNDI |
| | | | | RSYDRCKEGLNLSLWMDGGNVTKEBAIEAKGDFERAIRELLGLVQENTTIP |
| | | | | RAKDLFWKLMSIVNLVFMEDDGYTNRMLMNTVKAMFEPQMDLALLNK" |

Db 524 TGATCATATGACGAGGCCCTTGATTTTATCAAGTCTAATATAGCTTCAGCTACTGATGA 583
QY 635 AAGTCAACCACTCCCAITGGTTTGGATCATATATTTCTCGTGTGGTCTTGATGACGAA 594
Db 584 GAAACCAACGTTCTCCGGTGGGATTTGACATTAATTTTCCCTGGCATGATTTGATGCTAA 643
QY 695 AAACCTTGACATAAACCCTCTTTCAAAACAAACAGATTTTATGTTGATGCTACATAAGAG 754
Db 644 AGACTTGAATTTGAATCTACCCCTGGCACCGACGAACTGGATGCTTGGTTCGAAAGAA 703
QY 755 GGAATTTGAGCAAAAA-----AGATGCCATTTCAATGAGATGAGATGATCTTGGCGTA 808
Db 704 AGAGTTGGAGCTGAGAAGCTCGAGAGCAACTCTGAAGTGGAAAGCCTATTTAGCGTA 763
QY 809 TATCTCGAAGGACTCGGTAAATTTATATGATTTGGAATATGTTGAAGAAATATCATGAA 868
Db 764 TGTTCAGAAGGAATTTGAAAGTTTACAGGACTGGGATATGCTCATGCAATATCAAGGA 823
QY 869 AAATGGTTCTGTTTCACTCAACCATCAGCAAGCTGCTGCTTTTCAATATCATCAAAA 928
Db 824 GAATGGATCACTGTTTAAATTTCAATTCATCCACTACGCGAGCGCTTTTATGATAGAAATGA 883
QY 929 TCTGGTTGCTTAAATTTTAAATTTCACTTTTGGACAGTTTGGTAAATGAGTCCCAAC 988
Db 884 TGAATGCTGTTTGAATTTATCTTCTGCTCACTTTCAAAAAGTTTATGAGTCCAC 943
QY 989 AGTTTATCTCATGATTTATTTATCCGACTTTCTATGTTGACACAAATTTGAAGATTAGG 1048
Db 944 AATATATCTCTTGATATATATGCTCGATTACACATGTTGATAGCTTTCAAAAATTCGG 1003
QY 1049 AATTTCAACCAATTTGAGGTGGAATTTAAATTTTGAATTTTATGATGAAACATACAGATTTG 1108
Db 1004 AATTTGCTGGCATTTCAAGAGGAGATTTGAAGCGTATTAGATGAAACTTTACAGGTG 1063
QY 1109 GGTGGAACAGATGACGAAATTTATTCAGATTTTGAAGTGTGTAACATGCTTTAGCTTTGGTT 1168
Db 1064 GATGCAAGGAGAGGAAATATATTTCTTAGATGCTTCACTTTGCAATGGCTTTTGAAT 1123
QY 1169 ATTAAGGATCAATGGGTATGAAGTTTCCCGATCCATGCTGCTGAAATTAATA----- 1222
Db 1124 GTTACGTTTGAAGATATGATGTTCTTCAAGACAGTTGACTCAATTTTCAAGATAT 1183
QY 1223 -----TGAATTTAGCTTTGAAGACGATATGCTTCAAGACAGTTGACTCAATTTTCAAGATAT 1267
Db 1184 CTTTCCCAATGCTTTGGAGGATTTTAAAGACTTTGGGTGCTCGCTGGAGTTATATA 1243
QY 1268 TGCGTCAATATATATACCAAGAGATTTATCTTCT---GGAAAAACAATCTTGAAGTC 1324
Db 1244 GGCCTCTCAGATTTATCAAGCAACCCGATGAACTGTTCTGGAATAATATAACTCTTGGAC 1303
QY 1325 AGCTGATTTCCCAAGAGATATATPCCACTGATTTCAAAACAGGCTTTCTAAA----- 1376
Db 1304 TAGTCGTTTCCCTGAAGCATGGAATTTATCTAGTGAATTCAGTTTGGTCTGATAGAACGATAG 1363
QY 1377 -TTAATTTCAAAAGAGTGGAAAACTCTTAAGTTCCCTATCAATACCGTTTGAACG 1435
Db 1364 TGTCTTTAAACAGAGGCTGTTAATGCTCTTGAATTTCCCTATATAATGCAACTCTGAACG 1423
QY 1436 CATAAACACTAGACGAAATATACAGCTTTTCAATGTAGACAAATCAAGAAATTTCTGAAC 1495
Db 1424 CCTAATAAGTAAGAGGCAATGGAAGTTTACAGTGGAGACATTTGAGATTTCAAAATC 1483
QY 1496 TACATATCACTCAATATATAGTAAACATGATTAACCTTAAGTTGGCTTGTGAAGATTT 1555
Db 1484 GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGAACTTTGCTGATAGAGATTT 1543
QY 1556 CTACACCTGCCAATCTATTTATCGTGAAGAAATTTAAAGCTTTGAAAGTGGCTGTTAGA 1615
Db 1544 CAATACCTGCAACGCAATTTCACTTTAAGAACTGGAAGAGCTTTCAAGATGGGTGTTGA 1603
QY 1616 GAATAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCTCTGT 1675
Db 1604 AAACAAATTTGACGAGTTGAAATTTTTCAGACTGCACTGCACTAGGCTACTGTTTGTCTG 1663

QY 1676 TGCTCAACACATTTCTGCTCCGAATTTATCAGATCGGCTATTTTCAAGGCGCAAAATGG 1735
Db 1664 GGCAGCGACCTTTACTGATCTGAACTTTCATGATCTCGCATAGCATGGCACAATAGG 1723
QY 1736 CATATTAATCACTAGTATGATGATCTTTTGTATATCGGTGGTACAATCGATGAATGAC 1795
Db 1724 TGTGCTCAGCGCGGTGGTGGATTTCTATGATGGTGGAGGATCTGAAGAGGAATGGA 1783
QY 1796 CAACCTGATTTCAATGTTTGAATAATGGAATGTAGATCTGCACAGAGGATTTGTTTTCAGA 1855
Db 1784 TAACTTTATAGATTTGTTGGTGAAGAACTGGATCTGATGCGGGAAGTGGTTACTGTTCCAA 1843
QY 1856 GCATGTTCCGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGGATTTGGAGATGAAGC 1915
Db 1844 GGACGTTGAGATTTGATTTCTTGCACTGCACAGCACAGCTTTGTGAAATAGGAAGAGAGC 1903
QY 1916 TTTTAAATGGCAAGCGCGGATGTAACCTAGCCATGTTTATTAACCTGGTTGGATCTAAT 1975
Db 1904 TTTAGTATGGCAGGACGAGTGTATGAGGAATGTTATCGATGGTTGGTTGGCTCTGCT 1963
QY 1976 GAATAGTATGTTGAGAGAGCTATATGCAACAAGAGATGCTTTATGTCACCAATTAATGA 2035
Db 1964 GAAGGTGATGAGAAAGGAGCTGAATGTCGACAAATAAGGTAGTGCATCAATGGGTGA 2023
QY 2036 ATATATCGAAAAAGCTTTAGCTGATTTGATAGGCGCGGATTTCAAGCGCGCTATTTA 2095
Db 2024 ATATATGGAACAAGCCCATGATCATTTGCGGTTGGGACCTATAATCTTCCAATGCTCT 2083
QY 2096 CTTTGTGGGCGCCAAATTTATCAGAGGAGATTTGTAAGCTCTGAATATCATATCTATT 2155
Db 2084 CTTTGTGGACCTTAACTCTCAGAGGAATGATTTGGAAGCTGTGAATACCAGAGTTATA 2143
QY 2156 TAAGCTAATGAGCACGCGAGGTCGATCTTCAAAAGATATCCATAGCTTCAAGAGGGAAT 2215
Db 2144 TAAGCTGATGAGCACTGCTGCTGCTTGAAGATGATTTGATCTTACATAGAGATG 2203
QY 2216 TAAGGAGGCAATTAACGCGGTAGCATTTGCAATTTGAGTAAACGAGAAAGTGGGAAAGT 2275
Db 2204 CAAAGAGGAAAGCTGAATATTTCTGCTCTGTGGATGATTTGATGGGCTGATGTCAC 2263
QY 2276 GGAAGAGAGGTTGTCGAGGAGATGATGATGATTAATAAACAAGAGGAAAGATTAAT 2335
Db 2264 CAAAGAGGAGGCAATTTGAAGCAATTAAGGGATTTTGAAGGCGCATAGAGAGCTGT 2323
QY 2336 GAAATTAATTTTGAAGAAATTTGATAGCATTTGCTTAGAGCTTTGTAAGATGCAATTTG 2395
Db 2324 GGGGTTAGTTTTCAGAGGAAACACTACAA--TTCCAAGAGCTTTGAAGATTTGTTCTG 2380
QY 2396 GAACATGCTCAGCTGTTGAAATTTTTCAGCAACGATGACGGTTTACTGGAACAC 2455
Db 2381 GAAATGATGCTCAATTTGTAATCTATTTTACATGGAAGATGATGGGTACACTTCAATAG 2440
QY 2456 GATTTCTGATCTGTAAGGACATCATTTTCAACCCGTTGGTGGCTG 2502
Db 2441 GTTGTAGACACTGTAAGAGCCTGTTTGAACACCCATGGATCTGG 2487

RESULT 7
AF034774

LOCUS

2506 bp mRNA linear PLN 17-APR-1998

DEFINITION

Arabidopsis thaliana ent-kaurene synthase (GA2) mRNA, complete cds.

ACCESSION

AF034774.1

VERSION

AF034774.1

KEYWORDS

GI:3056724

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE

1 (bases 1 to 2506)

AUTHORS

Yamaguchi, S., Sun, T., Kawaide, H. and Kamiya, Y.

TITLE

The GA2 locus of Arabidopsis thaliana encodes ent-kaurene synthase

of gibberellin biosynthesis
Plant Physiol. 116 (4), 1271-1278 (1998)
98205064
9336043
2 (bases 1 to 2506)
Yanaguchi S., Sun, T.-P., Kawaide, H. and Kamiya, Y.
Direct Submission
Submitted (14-NOV-1997) Botany, Duke University, Research Dr.,
Durham, NC 27708, USA
Location/Qualifiers
1. .2506
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="1"
1. .2506
/gene="GA2"
68..2425
/gene="GA2"
/function="terpene cyclase"
/codon_start=1
/product="ent-kaurene synthase"
/protein_id="AAC39443.1"
/db_xref="GI:3056725"
/translation="MSINLRSSGSSPISATLERGLDSEVOTRANNVFEOTKEKIRK
MLEVEISVSAIDTSWAMVPSQONAPLPQCVKWLDDNOHEDGSLNDHDSOIL
KQVLSSTLALIKWIGIGERQINKLOFIELNSALVTDETIQKPGFDIIPGMI
KYARDLNLITPLGSEVDDMRKRDLDKCDSEKSEKREAYLAVLEGTRNLKDWL
IVKYORKNSLFDSPATAAFTQPGNDGCLRYLCSLQKFEAAVPSVYPPDVARLS
LIVLESGLIDRDKTEIKSLIDETRYWLRGDBEICLDLATCALAFLLLAHGYDS
YDPKPAEBSGFSDTLEIGYVNTFSVLELFKAAQSPHESALKKQCWTQYLEMEL
SSWKTYSVRDKLKEVEDALAFPSYASLERSDHRKILNGSAVENTRVKTSVRLHN
ICTSDILKALDFNFCQSIHREMERLDRIVENRLOELKFAKOKLAYCVFSGAATL
FSPSLDARISWAGGVLTVVDDPFDVGGSKLEHLHLHVEKWDINGVPEYSEHV
EIIIFSVLRDITLEFGKAFYQGRNVTHHIVKWLMLKREARESSDKSPSLSD
YMENAYISFALGPVLPATYILGPPLEKTPVDHSHQYNOLYKLVSTMGRLNDIQGPKR
ESAEKGNLNSLHMKHEDNRKSEVITESMKGLAERKRREBLHLKLVLEKSGSVPRECK
EAFLLMSKVLNLFYRKDDGFTSNDLSLVKSIVPEVPSLOKESLT"

Query Match 26.2%; Score 732.4; DB 8; Length 2506;
Best Local Similarity 59.5%; Pred. No. 2.5e-133;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

QY 270 GGACAAACTAATCCCACTAATCTGATCATATGATACCAACCAAGAACGATCCAAACACAG 329
DB 143 GTACAGACAAGAGCTAACATGTCAGCTTTGAGCAACAAGAGAGAGATTAGGAAGATG 202
QY 330 TTTAAATGTAGAAATTTCTGTTCTTCATATGACACAGCATGGGTAGCCATGGTCCCT 389
DB 203 TTGGAGAAAGTGGAGCTTCTGTTTCGGCCTACGATACCTAGTTGGGTAGCAATGGTTCCA 262
QY 390 TCTCCAAACTCACCAATCCCTGTTGTTCCCTGAGTGTCTCAATTGGTTAAATAAAT 449
DB 263 TCACCGAGCTCCCAAAATGTCACCTTTTCCACAGTGTGGAATGGTTATTGGATAAT 322
QY 450 CAGCTTAATGATGGTTTCATGGGCTTCTGTTAAATCACACTCATATCATATCAACCCGCTG 509
DB 323 CAACATGAAGATGATCTTGGGACTTGATAACC-----ATGACCAATCAATCTCTT 373
QY 510 CTTAAAGATTCTCTAATCTCAACATTAGCANGATTGTTGTCATTAATAAAGATGGAATGTT 569
DB 374 AAGAGAGATTGTTATCATCTACACTGGCTAGTATCTCGCGTTAAAGAAGTGGGAATT 433
QY 570 GGGGAAGATCAAAATAAATGAGTCTAAGTTTATTGAGTCAAAATCTTGCTTCACTACT 629
DB 434 GGTGAAGACAAATAAACAAGGCTCCAGTTTATTGAGCTGAATCTGCAATTAGTCACT 493
QY 630 GAAAAAAGTCAACCATCTCCCATTTGGTTTGTGACATCATATTTCTGTTGTTGTTGAGTAT 689
DB 494 GATGAACCAATACAGAAACCAACAGGGTTGTGATTTATTTCTCTGGGATGATTAATAT 553
QY 690 GCGAAAAACTTGGACATAAACCTCTCTTTCAAAACAAACAGATTTTAGTTGATGCTACAT 749

ORIGIN

DB 554 GCTAGAGATTTGAATCTGACGATTCATTTGGGCTCAGAAAGTGGTGGATGACATGATACGA 613
QY 750 AAGAGGGAATTCGAGCAAAAAAG-----ATGCCATTCAAAATGAGATGAGATGATAC 800
DB 614 AAAAGAGATCTGGAATCTTAAATGTGATAGTGAAGATTTTCAAAGGGAAGAGAGATAT 673
QY 801 TTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGTTGAGAAATAT 860
DB 674 CTGGCCTATGTTTATAGAGGGGACAAAGAAACCTTAAAGATTGGGATTGATAGTCAATAT 733
QY 861 CAGATGAAAAATGGTTCTGTTTCACTCACCATCAGCAACAGCTGCTGCTTCATTAAT 920
DB 734 CAAAGGAAAAATGGGTCTACCTGTTGATTCACGACCAACACAGCAGCTGCTTTACTCAG 793
QY 921 CATCAAAATCCTGGTGTCTTAAATTTAAATTTCACTTTTGGACAAGTTTGGTAATGCA 980
DB 794 TTTGGGAATGATGTTGTTCTCCGTTATCTCTGTTCTCTCTTTCAGAAATTCAGGCTGCA 853
QY 981 GTCCCAACAGTTTATCTCTCATGATTTATTATTCGACTTTCTATGTTTGACAATTTGAA 1040
DB 854 GTTCTCTCAGTTTATCCATTTGATCAATATGCAACGCTTAGTATATATTTGCTCACTCTTGA 913
QY 1041 AGATTAGGAATTTCAACACCATTTGAGTGGAAATTAATAATGTTTATAGTGAACATATC 1100
DB 914 AGCTTAGGAATTTGATAGATTTTCAAAACCGAAATCAAAAGCATATTTGGATGAACCTAT 973
QY 1101 AGATTGGGTGGAAACAGATGAGCAATATATTCATGATGTTTAAACATGCTTTGATGCC 1160
DB 974 AGATATTGGCTTCCTGGGGATGAAGAAATATGTTTGGATTTGGCCACTTTGCTTGGCT 1033
QY 1161 TTTGGTTATTAAGGATCAATGGGTATGAAGTTTCCCAGATCCATTTGGCTGGAATTTACT 1220
DB 1034 TTCCGATTTATGCTTGTCTCATGGCTGATGATGTTTACGATCCGTAAACCACTTTGCA 1093
QY 1221 AATCAATTAGCTTTGAAAGACGATATGCGACTCTTGAACATATCATGCGTCAATATA 1280
DB 1094 GAAAGATCTGGTTCTCTGATCTTTTGAAGGATGATGTTGAAGATAAGTTTCTGTGTTA 1153
QY 1281 TTATACCAAGAGAT-----TTATCTTCTGGAAAAAATAATCTTGAAGTCACTGATTTCC 1335
DB 1154 GAATTTATTTAAGGCTGCTCAAGATTATCCATGAATCAGCTTTTGAAGAGCAGTGTGT 1213
QY 1336 TCAAGAGATTAATTCACCTGATTCAA-----ACAGGCTTTCT 1373
DB 1214 TGGACTAAACATATCTGGAGATGGAATTTCCAGCTGGGTTAAGACCTCTGTTCCGAGAT 1273
QY 1374 AAATTAATTCACAAAGAGTGGAAATGCTCTTAAAGTTCCCTATCAATACCGGTTTGA 1433
DB 1274 AAATACCTCAAGAAAGAGTGGAGATGCTTCTTGTCTTTCCTCTATGCAAGCCTAGAA 1333
QY 1434 CGCATAAACTAGACGAAATATATAC---AGCTTTTCAATGTAGACAATACAAGATTTCTG 1490
DB 1334 AGATCAGATCACAGGAGAAAAATACTCAATGGTTCTGCTGTGGAACACACAGATTACA 1393
QY 1491 AAAACTCATATCATCATCAATAATTTAGTAACATGATTAACCTAAGTTTGGCTGTTGAA 1550
DB 1394 AAAACCTCATATCGTTTGCACATATTTGCACCTCTGATATCTCTGAAGTTAGCTGGAT 1453
QY 1551 GATTTCTACACCTGCGCAATCTATTTATCGTGAAGATTAAGGCTCTTGAAGGTTGGGT 1610
DB 1454 GACTTCAATTTCTGCGAGTCCATACACCGTGAAGAAATGGAACGTCTTGATAGTGGATT 1513
QY 1611 GTAGAGAAATAGTTGGACCGAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTTC 1670
DB 1514 GTGAGAAATAGATTGACGAACTGAATTTGACAGACAGAGCTGGCTTACTGTTATTTTC 1573
QY 1671 TCTGTTGCTGCAACATTTCTGCTCCCGAATTAATCAGATGCGGTTATTTTCATGGGCCAA 1730
DB 1574 TCTGGGGCTGCAACTTATTTTCTCCAGAACTATCTGATGCTCGTATATCGTGGGCCAA 1633
QY 1731 AATGGCATATTAACATACAGTACTTGTGATGATTTTGTGATATCGTGGTGAATTCGATGAA 1790

| | | | | |
|------------|--|---|-----------------|---|
| Db | 1634 | GGTGGAGTACTTACAAACGGTTGTAGACGACTTCTTTTGATGTTGGAGGTCACAAAGAGAA | 1693 | Upadhyaya,N.M. Direct Submission Submitted (21-JUL-2003) Rice Functional Genomics Group, Genomics and Plant Development Program, CSIRO Plant Industry, Chr. Barry Drive and Clunies Ross Street, GPO Box 1600, Canberra, ACT 2601, Australia |
| Qy | 1791 | TTGACCAACCTGATTCAAATGTTGAAATATGGAATGTAGATGTGCGACAGATGTTGCT | 1850 | |
| Db | 1694 | CTGGAACACCTCATACACTTGGTCGAAAGTGGGATTTGAACGGGTGTTCTCTGAGTACAGC | 1753 | |
| Qy | 1851 | TCAGAGCATGTTCGGATTTTATTTTACCATTAATAAGATGCAATCTGTTGGATTTGGAGAT | 1910 | |
| Db | 1754 | TCAGAACATGTTGAGATCATATTCCTCAGTTCTTAGGGACACATCTCGAACAGAGAC | 1813 | |
| Qy | 1911 | GAAGCTTTTAAATGCGACGCGGCAATCTAATAGCCATGTTTATTCAACTTGGTTGAA | 1970 | |
| Db | 1814 | AAAGCATTCACCTATCAAGAGCGCAATGTGACACACCAATTTGTGAAATTTGGTTGGAT | 1873 | |
| Qy | 1971 | CTAATGAATAGTATGTTGAGAGACCTATATATGGAACAAGATGCTTATGTGCGCAATTA | 2030 | |
| Db | 1874 | CTGCTCAAGTCTATGTTGAGAGAACCGGAGTGGTCCAGTGAAGTCAACCAAGCTTG | 1933 | |
| Qy | 2031 | AATGAATATATGGAACCGCTTACCTGTCATTTGATTAGGCGCCGATTTGTCGAAGCGCT | 2090 | |
| Db | 1934 | GAGGATTCATGGAATAACCGTACATATCATTTGATTAAGGCCAATTTGCTCCAGCT | 1993 | |
| Qy | 2091 | ATTTACTTTTGGGGCCCCAAATATCAGAGGAGATTTGTGAAGCTCTGAATATCATAT | 2150 | |
| Db | 1994 | ACCTATCTGATCGGACCTCCACTTCCAGAGAGACAGTGTGATGCCAATATAATCAG | 2053 | |
| Qy | 2151 | CTATTAAGTAAATGAGCAGCGGTCGACTTCTTAACGATATTCATAGCTTCAAGG | 2210 | |
| Db | 2054 | CTCTACAAGCTCGTAGCCTATGGTCTCTTCTTAATGATACATCAAGGTTTAAAGAA | 2113 | |
| Qy | 2211 | GAATTTAAGGAAGGCAATTAACCGGTTAGCATTTGATTAACCGAGAAAGTGGG | 2270 | |
| Db | 2114 | GAAGCGCGGAAGGAGCTGAATGCGGTTTCAATGCAATGAACACGAGAGACAT | 2173 | |
| Qy | 2271 | AAAGTGAAGAGAGAGTGTGGAGGAGATCATGATGATTAATAAACAGAGGAAGAA | 2330 | |
| Db | 2174 | CGCAGCAAGAGTGAATAGATCGATGAAAGTTTACGACAGAGAAAGGGAAGAA | 2233 | |
| Qy | 2331 | TTAATGAATTAATTTTGAAGAAATGTTAGCATTTGTTCTTAGACITTTGAAGATGCA | 2390 | |
| Db | 2234 | TTGCAATAGTATGTTTGGAGGAGAAAGGAGTGTGTTCCAAAGGAATGCAAGAGCG | 2293 | |
| Qy | 2391 | TTTGTGAACATGTCACGTTGTTGAATTTTTCAGCAACCAAGATCAGCGGTTTACTGGA | 2450 | |
| Db | 2294 | TTCTTGAATGAAGCAAGTGTGAATTTTATTTTACAGAGGACGATGATCATCA | 2353 | |
| Qy | 2451 | AACACGATTTTGATCTGTGAAGGACATCATTTTCAACCCGTTGTTGTTGAATGAA | 2510 | |
| Db | 2354 | AATGATCTGATGATCTTGTGTTAAATCAGTGTATCTACGAGCTGTAGCTTACAGAAAGAA | 2413 | |
| RESULT 8 | AY347876 | 2271 bp mRNA linear | PLN 05-AUG-2004 | |
| LOCUS | AY347876 | | | |
| DEFINITION | Oryza sativa (japonica cultivar-group) map BAC OSJNBa0070C17 | | | |
| ACCESSION | AY347876 | | | |
| VERSION | AY347876.1 | GI:34452213 | | |
| KEYWORDS | | | | |
| SOURCE | Oryza sativa (japonica cultivar-group) | | | |
| ORGANISM | Oryza sativa (japonica cultivar-group) | | | |
| REFERENCE | 1 (bases 1 to 2271) | | | |
| AUTHORS | Upadhyaya,N.M. | | | |
| TITLE | Isolation and characterization of a Ds-tagged rice (Oryza sativa L.) GA-responsive dwarf mutant defective in an early step of the gibberellin biosynthesis pathway | | | |
| JOURNAL | Unpublished | | | |
| REFERENCE | 2 (bases 1 to 2271) | | | |
| AUTHORS | Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,B.S. and | | | |

| | | | | |
|-----------------------|--------------|---|---|--------------|
| misc_difference | 1087 | /gene="OsKSI1A" | /note="compared to genomic sequence in GenBank Accession Number AL731610" | |
| ORIGIN | | /replace="t" | | |
| Query Match | 20.1%; | Score 559.8; | DB 8; | Length 2271; |
| Best Local Similarity | 56.0%; | Pred. No. 1.7e-99; | | |
| Matches 1215; | Conservative | 0; | Mismatches 902; | Indels 54; |
| Gaps | 6; | | | |
| Qy | 308 | CAAAGAACGATGCCAAAAACAGTTTAAAAATGTAGAAAATTTCTGTTTCTTCATATGACAC | 367 | |
| Db | 21 | CCAGGCAAGACAAAGGACCGAGCTCCAGACATTTGGAGTTATCTACGCTTTGTACGACAC | 80 | |
| Qy | 368 | AGATGGGTAGCATGTCCTTCTCAAACTCACCACATCGCTTGTTCCTCGATGAGT | 427 | |
| Db | 81 | GGCGTGGTGGCCATGTCGCACTGCGGGGTCCCGTCAGCATCCGCTTCCCAATG | 140 | |
| Qy | 428 | TCCTCAATGCTGTTAAATTAATTAATCAGCTTAAATGATGTTTATGTTGTTCTTCTTAATCACAC | 487 | |
| Db | 141 | CGTGGAGTGAATTAATTAACAAACGAGGATGATGATCTTGGGT | 188 | |
| Qy | 488 | TCATAATCAATATCACCCTGTTTAAAGATTTCTTATCTTCAACATAGCATGATTGT | 547 | |
| Db | 189 | AAGAGGATTCGGCGTGGCGGTCAAGAGAGATTTCTGCTCTACGTTGCGTGTGTCT | 248 | |
| Qy | 548 | TGATTAATAAGATGGAATCTTGGGAGAGATCAATAATAAAGGCTTAAGTTTATTTGA | 607 | |
| Db | 249 | TGCATCAAGAGATGGAATGTTGGGAGGAGACATCAGGAGGAGACTAGATTATTG | 308 | |
| Qy | 608 | GTCAATCTTCTGCTTACGCTACTGAAATAAGTCAACCATCTCCATTTGGTTTGAATCAT | 667 | |
| Db | 309 | AAGGAATTTCTCCATTTCCATGGATGAGCAGATTTGCTCTCTGTTTCAACATCAC | 368 | |

QY 668 ATTTCCTGGTTTGGCTTGAATATGCGAATAAATTGGACATATAACCTCCCTTTCACAAACAAC 727
Db 369 TTTTCTCGTATGCTTGAACCTTGCATGGGATGGAATTAGATTCCCTGTGACAGAC 428
QY 728 AGATTTTATGTTGATGCTACATAGAGGGAATTGAGCAAAAAA-----GATGCCA 778
Db 429 GGAATGTTGATAGACTTCTTCCCTCCGGAGATAGAACTGGAAAGAGAGGCTGGAGACCA 488
QY 779 TTCAAATGAGATGAGATGATGCTTGGCTATATCTCTGAAGACTCGTAAATTTATATGA 838
Db 489 TTTCTATGGAAGAAAGGCAATATGCTTATGCTTACAGAGGATTAGGAAACCTGTTGGA 548
QY 839 TTGGAATATGTTGGAAGAAATATCAGATGAAAAATGTTCTGTTTCAACTCACCATCAGC 898
Db 549 GTGGATCAGATATGATGTTTCCAGAGGAAGATGGAATCAITTTTCAACTGCTCTTCCAC 608
QY 899 AACAGCTGCTCTTTCATTAATATCAAAAAATCCTGGTTGCTTAAATTTAAATTCAC 958
Db 609 AACTGCTGCCACATAGTCAACCACTATAACGATAAAGCCCTCCAGTACTTAAATTTGCT 668
QY 959 TTTGGACAAGTTTGGTATGATGAGTCCCAACAGTTTATCTCTCATGATTTATTTATCCGACT 1018
Db 669 GGTAGTAAATTTGGCAGTGCAGTACCAACAGTGTATCCATTAACATATATGTCAGCT 728
QY 1019 TTTCTATGTTGACCAATTTGAAGATTAGGAATTTTCAACCAATTTTCAAGTGGAAATTA 1078
Db 729 TTTATGGTGTGATGCTTGGAAAGATGGGAATTTCTCAGTATTTTGTCTAGTGAGATAA 788
QY 1079 AAATGTTTATGATGAACATACAGATGTTGGGTGGAACGAGATGAGCAATATTCATGGA 1138
Db 789 GAGATCCTAGACACAACATCGTTTCATGTTTAGAAGAGATGAGGAAATTTATGCTGA 848
QY 1139 TGTGTTAACTATGCTTTAGCTTTGCTTATTAAGGATCAATCGGTATGAGTTTCCCC 1198
Db 849 CATTAACAATGTCATGCTTTTCCCTGTTAAGGATGATGGGTATCATGTTTCTCTC 908
QY 1199 AGATCCATTTGGCTGAATTAATA-----TGAATTAGCTTT 1234
Db 909 AGTCGAGTATCTCTGCTGTAAGCTTCCAGTTTCCGTGAATCACTTCAAGATATTT 968
QY 1235 GAAAGCAATATGCACTCTTGAACATATCATGCTGCAATATATATACCAAGAGGA 1294
Db 969 AAATGATAGAAATCTCTAAATGCAATATATACAGCTTCAAAAGTCAATTAATCAGAAA 1028
QY 1295 TTTATCTTCTGGAACAAATCTTGAAGTCAAGTCTGATTTCTCTCAAGATATATCCAC 1354
Db 1029 TGAATCAATCTTAGATAGCATAGCTCTTGGTCAGGTAGCTTATTTGAAGAAAGTGTGAG 1088
QY 1355 TGATTCAAAAGGCTTTCTAAATTTAAATTCAAAAGAGGTGGAATAATGCTCTTAAGTTCCC 1414
Db 1089 CTCTAATGGGCTGAAAAAAGCTCCAAATCTTGAAGAGATGAATGCTCTTAAGTTTCC 1148
QY 1415 TATCAATACCGGTTTAGAAGCTATACACTAGACGAATATACAGCTTTACAAATGAGA 1474
Db 1149 CTCTACACCACTGGACCGCTTAGATCAACAGAGGAACAATTTGATGCAAA 1208
QY 1475 CAATACAAGAAATCTGAAAAATCATATCACTCATCAAAATATTAGTAACACTGATTA 1534
Db 1209 GGATTTCCAGATTTAAGACGAAT---ACTTGTCTCTCATGCCATCAAGATATCT 1265
QY 1535 AAGTTGGCTGTTGAAGATTTTACACCTGCCAATCTATTATTCGTGAAGAAATTAAGG 1594
Db 1266 AGCTTTGGCTGTTGAAGATTTAGTAGTTCTCAATCTATATACAGATGAATTAATTA 1325
QY 1595 TCTTGAAGGTTGGGTAGAGATAGTTGGACAGCTCAAGTTGCTAGGCAAAAGAC 1654
Db 1326 TCTTGAAGTTGGGTGAAGATGAAAAGCTTCAGTCACTGCCATGCCAGCAGAGTT 1385
QY 1655 CGCTACTGTTATTTCTCTGCTGCAACACATTTGCTCTCCGGAATATCAAGTGGG 1714
Db 1386 GACATATGCTACTTTCTGCTGCTGCTACATATTTCCCGGGAATGCTGAAGCCCG 1445

QY 1715 TATTTTCATGGCCAAAATGGCATATTAACCTACAGTAGTGTGATGACTTTTGTGATATCGG 1774
Db 1446 CATTTGCATGGGCTAAAAATGGTGTACTGACCAACTGTTGTTGATGACTTCTTTGATCTTGG 1505
QY 1775 TGTACCAATCGATGAATTTGACCAACCTGATTCAAATGCTTCAAAAATCGAATGTAGATGT 1834
Db 1506 GGGATCAAAAGAGNACTAGAAACCTCATTTGCTTTAGTAGAAGUGGATGGACAT-- 1563
QY 1835 CGACAAGATTTGTTGTTTCAGAGCATGTTTCGATTTATTTTATTTAGCAATTAAGATGCAAT 1894
Db 1564 -CAAGAGGATTTCTACTCAGAACCAAGTAAGATAGTTTTTCTGCTATTTATCTACAGT 1622
QY 1895 CTGTTGATTTGAGATGAAGCTTTTAAATGCAAGCGCGCATGTAACTAGCCATGTTAT 1954
Db 1623 GAACAGCTTTGGGCAAAAGGCTTCTGCAATTAAGAGCGCTGATGTTACAAAACCACTAAC 1682
QY 1955 TCAAACTTGGTTGGAACCTAATGAATATGATGTTGAGAGCAAGCTATATGACAAAGATGC 2014
Db 1683 AGAAATATGTTTATGTTCTGATGAGGTCTATGATGCCGAAGCTGATGCGAGAGACAA 1742
QY 2015 TTATGTGCCAACATTAATGAATATATGAAACGCTTACCTGTCATTTGCAATTTAGGCC 2074
Db 1743 ATATGTCCGCAATGGAAGAAATATATGGCAATATGCTGTTGCTCTCAITTCACCTGGAC 1802
QY 2075 GATTTCAAGCGGCTATTTACTTTTGGGGCCCAAAATTTATCAGAGGATTTGTTGAAG 2134
Db 1803 TATTTGCTCTCCGACTCTGTTATTTTCGAGGACCAAGCTTCCAGAGGATGTCGTACGGGA 1862
QY 2135 CTCGTAATATCATTAATCTATTTAAGCTAATGAGCACCGGCTCGACTTTTAAACGATAT 2194
Db 1863 TCATGAGTCAATGAATTAATTTAGCTATGAGCACTTTGTTGGGCTCTCTGTAATGACAG 1922
QY 2195 CCATAGCTTCAAGAGGGAATTTAAGGAAGCAATTAACCGGTAGCATTTGCAATTTGAG 2254
Db 1923 CCAAGGCTTTGAGAGGGAGAGCCCTGGAGGGAAGCTGAACAGTGTCTCACTGCTTTGTTCA 1982
QY 2255 TAACGAGAAAGTGGGAAAGTGGGAAGAGGTTGTGGAGAGATGATGATGATTA 2314
Db 1983 TCAAGTGTGTTTCTATCTCCATAGACGAGGCTTAAATGAAGCCCAAGAAATCCATAGA 2042
QY 2315 AAACAAGAGAAAGAAATTAATGAATTAATTTTGAAGAAATGTTAGCATTTGTTCTAG 2374
Db 2043 CACTTCCAGGAGGAACCTTGTAGATTTGCTTGGAGAACAGGT---GCTGTTCTTAG 2099
QY 2375 AGCTTGTAAAGATGCAATTTTGGAAACAGTGTACGTTTGAATTTTTCACGCAACGA 2434
Db 2100 GCCATGCAAGCAGCTGTTCTGGAAGATGTGCAAGATTTGTTCAATGTTCTACTCTCGGAC 2159
QY 2435 TGACGGGTTTA 2445
Db 2160 TGATGGGTTTA 2170

RESULT 9
AV347877 2283 bp mRNA linear PLN 05-AUG-2004
LOCUS Oryza sativa (japonica cultivar-group) clone OsKS-mutant clone 1
DEFINITION ent-kaurene synthase 1B (OsKS1B) mRNA, complete cds.
ACCESSION AY347877
VERSION AY347877.1 GI:34452215
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 2283)
AUTHORS Margis-Ponheiro, M., Zhou, X.-R., Zhu, Q.-H., Dennis, E.S. and Upadhyaya, N.M.
TITLE Isolation and characterization of a Ds-tagged rice (Oryza sativa L.) GA-responsive dwarf mutant defective in an early step of the gibberellin biosynthesis pathway
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2283)
AUTHORS Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,E.S. and Upadhyaya,N.M.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2003) Rice Functional Genomics Group, Genomics and Plant Development Program, CSIRO Plant Industry, Cnr. Barry Drive and Clunies Ross Street, GPO Box 1600, Canberra, ACT 2601, Australia

FEATURES
source Location/Qualifiers
1..2283 /organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
clone="OsKS-mutant clone 1"
1..2283 /gene="OsKS1B"
1..2271 /gene="OsKS1B"
/note="ent-kaurene synthase 1-like protein OsKS1B; similar to OsKS1A sequence in GenBank Accession Number AY347876; similar clones obtained from wild type and dwarf mutant TT14-1-3-1-1 plants"
/codon_start=1
/product="ent-kaurene synthase 1B"
/protein_id="AAQ72560.1"
/translation="MOHKKELQARTRDQLELSTSIYDIAWVMPILRSRQPCF PQCVWIIQODDSWGTGFGVATRDVLSLACVLAALRWNVGGEHRRGLDFI GNFISAMDEQIAAPVFNITPFGMLSLAMGDMLEFPVRQTDVRLHLRELEPERA GDHSYGRKAYMAVYVPLNTYCOLSWDALEKMGISQYFVSEIKSILDITVVSWLER YLNCVLFKFSNAPVYVPLNTYCOLSWDALEKMGISQYFVSEIKSILDITVVSWLER DEIMDLITTCAMARLLRMNGYHVSSELSPVAESPFRESLQYVLDNKKSLIELYK ASKVSSENLSDLSIGSWGLLESVSNVKKAPIFEEMKIALKPFYITLRLDL AKRNEREDAKDSQMLKTGYLLPHANQDILALAVEDFSSQSIYODENLNLECWKDE KLDQLPSARQKLYCVLSAATIFPRLSEARIAWAKNGVLTIVVDVDFDLGGSKBEL ENLIALVKWMDHOFBEFYSEQVRIIVFSAITVGTOLGAKASALQDRDVFNLTEWLC LMSWMAEWMQRTKVPTMEESYMANAVSPALGPVLPVLFVPGKQEDVVRDHEY NELFRIMSTCGRLNDSQGFERESLEGLNSVILVHSGSISIDEAKMAKQKSIDT SRNLLRLVLGEGQAVPRCPKQLFWKMKLVENFYRDTGDFSPKEMSVANVVKPEP LKLRVSDPYGSLISGN"

ORIGIN
Query Match 19.9%; Score 555; DB 8; Length 2283;
Best Local Similarity 55.8%; Pred. No. 1.5e-98;
Matches 1212; Conservative 0; Mismatches 905; Indels 54; Gaps 6;
308 CAAAGAACGGATCCAAAAACAGTTTAAAAATGTAGAAATTTCTGTCTTCTCATATGACAC 367
Db 21 CCAGGCAAGGACAGGACAGCTCCAGACATTTGGAGTTATCTACGCTCTTTGACGACAC 80
368 AGCATGGGTAGCATGGTCCCTTCTCAAACTCACCACCAATCGCCTGTGTTCCCTGAGTG 427
Db 81 GCGGTGGGTGGCCATGGTGCACCTGCGGGGTTCCTCGTCAGCATCCGCTGTTCCCAATG 140
428 TCTCAATGCTTAATTAATAATCAGCTTAATGATGTTTCAATGGGCTCTTGTAAATCACAC 487
Db 141 CGTGGAGTGATTTACAAACCAAGCAGGATGATGATCTTGGGGT-----AC 188
488 TCATAATCATATATCAACCCGTTGCTTAAAGATTTCTATCTTCAACATTAAGATGATGTT 547
Db 189 AAGAGGATTCGGGTGGCGGTCAACAGAGATGTTCTGCTACGTTGGCGGTGTGTTCT 248
548 TGCAATTAAGATGAGATGTTGGGAGAGATCAATTAATAAAGAGTCTAAGTTTATTGA 607
Db 249 TGCACATCAAGATGGAATTTGGGAGGAGGACATCAGAGAGGACTAGATTTATTTGG 308
608 GTCAATCTTGTCTCAGTACTGAAAAAGTCAACCATCTCCCATTTGGTTTGGACATCAT 667
Db 309 AAGGAATTTCTCAATGCCATGATGAGCAGATGTGCTCTCTGTTGGTTTCAACATCAC 368
668 ATTTCTCTGGTTGCTTGGAGTATGCGAAAAACTTGGACATAAACCTCTTTCAAAACAAAC 727
|||||

Db 369 TTTTCTCTGGTATGCTTAGCCTTGCCATGGGATGGATTTAGAAATTCCTCTGTCAACAGAC 428
Qy 728 AGATTTTAGTTTGTATGCTACATAAGAGGCAATTCGAGCAAAAA-----GATGCCA 778
Db 429 GGATGTTGATAGACTTCTTACCTCCGGAGATAGAACTGGAAGAGAGGCTGGAGACCA 488
Qy 779 TTCAAAATGAGATGATGATGATCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATGA 838
Db 489 TTTCTATGGAAGAAGGCATATATGCTTATGTCACAGAAGATTTAGGAAACCTGTTGGA 548
Qy 839 TTGGAATATGTTGAAGAAATATCAGATCAAAATGGTCTGTTTCAACTCACCATCAGC 898
Db 549 GTGGGATGAGATTTATGATGTTCCAGAGGAAGATGGATCATTTTCAACTGCTCTCCAC 608
Qy 899 AACGCTGCTGCTTTTCAATTAATCATCAAAATCTCGTGTCTTAAATTTTAAATTCAT 958
Db 609 AACTGCTGCCACATTTAGTCAACCACTATAACGATAAAGCCCTCCAGTACTTAAATTTGCT 668
Qy 959 TTTGACAAGTTTGGTAAATGAGTCCCAACAGTTTATCTCTCATGATTTATTTATCCGACT 1018
Db 669 GGTTAGTAAATTTGGCAGTGCAGTACCAACAGTGTATCCACTAAACATATATTGTGACT 728
Qy 1019 TTCTATGCTTGACACAATTTGAAGATTAGGAATTTACACCACTTTACAGAGTGGAAATTA 1078
Db 729 TTTATGGGTGGATGCACTTGAAAAGATGGGAATTTCTCAGTATTTTGTCAAGTGAATAA 788
Qy 1079 AAATGCTTTTATGATGAACATACAGATGTTGGTGGAAAGAGATGAGCAAAATTTATGGA 1138
Db 789 GAGCACTTAGACACACATAGCTTTTATGTTAGGAAGAGATGAGGAATTTATGCTGA 848
Qy 1139 TGTCTTAACATGCTTTTAGCCTTTTGGTATTAAAGATCAATGGGTATGAAGTTTCCCT 1198
Db 849 CATAACAACATGCAATGGCTTTTTCGCTGTTTAAAGATGAATGGGTATCATGTTTCTCTC 908
Qy 1199 AGATCAATGGCTGAAATTAATA-----TGAATTAGCTTT 1234
Db 909 AGTCAGATTTATCTCTGTTGCTGGAAGCTTCAGATTTCCGTGAATCACTTCAAGGATATT 968
Qy 1235 GAAAGACGAATATGACAGCTTTTGAACATATCATCGTGCACATATATATATACCAAGAGA 1294
Db 969 AAATGATAAGAAATCTTAATCGAATTATCAAGGCTTCAAAAGTCAGTAAATCAGAAA 1028
Qy 1295 TTTATCTTCTGGAAACAAATCTTGAAGTCAGCTGATTTCTTCAAGAGATATATATCCAC 1354
Db 1029 TGAATCAATCTTAGATAGCATAGGCTCTTGGTCAGGTAGCTTATTGAAGGAAAGTGTAG 1088
Qy 1355 TGAATCAACAGCTTTTAAATTAATTCACAAAGAGGTGGAAATGCTCTTAAAGTTCC 1414
Db 1089 CTCATAGGGGTGAAAAAGCTCCAAATCTTTGAAGAGATGAAGTATGCTCTTAAAGTTCC 1148
Qy 1415 TATCAATACCGGTTTAGAACGGCATAAACCTAGACGAAATATACAGCTTTTACAATGTAGA 1474
Db 1149 CTTCTACACCACTGGACCGTCTAGATCAAGAGGAACATTTGAACGTTTGTATGACAA 1208
Qy 1475 CAATACAGAAATTTGAAACCTACATATCACTCATCAATATATAGTAACACTGATTAACCT 1534
Db 1209 GGAATCCAGATGTTTAAAGACGGGAT---ACTTGTCTCTCATGCCAATCAAGATATCT 1265
Qy 1535 AAGGTTGGCTGTTTGAAGATTTCTTACACCTGCAATCTTATTTATCTGTAAGAAATTAAG 1594
Db 1266 AGCTTTGGCTTTGAAGATTTTCACTAGTTTCTCAATCTATATACAGGATGAATTAATA 1325
Qy 1595 TCTTTGAAGGTTGGGTAGAGAAATAGTTGGACAGCTCAAGTTTGTAGGCAAAAGAC 1654
Db 1326 TCTTTGAGTGTGGGTGAAAGATGAAAGCTCGATCAGCTGCCATCTGCGAGGAGGAT 1385
Qy 1655 CGCTTACTGTTTCTCTGTTGCTGCAACACTTTCTGTTCCCGAATATCATAGTCGCG 1714
Db 1386 GACATATGCTACCTTTCTGCTGCTGCTACCATTAATTTCCCGCTGGAATTTCTGAGGCCG 1445
Qy 1715 TATTTTCATGGCCCAAAATAGGCATATTAACATCAGTAGTTGATGACTTTTGTGATFCGG 1774
Db 1446 CATTCATGGGCTTAAAAATGGTGTACTGCAACTGTGTTGATGACTTCTTTGATCTGG 1505

Db 429 GGATGTTGATAGACTTCTTCCCTCCGGAGATAGAACTGGAAAGAGAGGCTGGAGACCA 488
 Qy 779 TTCAAATGAGATGAGATGATCTTGGCGTATATCTCTGAAAGGACTCGGTATATATATGA 838
 Db 489 TTCTTATGGAAGAAAGGATATATGCTTATGTCACAGAGGATTAGGAACCCGGTTGGA 548
 Qy 839 TTGGAATATGTTGAAGAAATATCATGATGAAAGAAAGTCTGTTTCAACTCACCATCAGC 898
 Db 549 GTGGGATGAGATATGATGTTCCAGGAGAAAGATGATCAATTTTCAACTGTCCTTCCAC 608
 Qy 899 AACAGCTCTGCTTTCATTAATCATCAAAATCCTGTTGCTTATTAATTTAAATTCAT 958
 Db 609 AACTGCTGCCATATGATGTTCAACCACTATAACGATAAAGCCCTCCAGTACTTAAATTTGCT 668
 Qy 959 TTTGACAAAGTTTGTGTAATGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACT 1018
 Db 669 GGTAGTAAATTTGGCAGTGCACTTGAAGATGGAATTTCTCAGTATTTTGTGTCAGTGATAA 728
 Qy 1019 TTCTATGTTGACACAATTTGAAGATTTAGGAATTTTCAACCAATTTTCAAGAGTGAATTA 1078
 Db 729 TTTATGGTGGATGACCTTTGAAGATGGAATTTTCTCAGTATTTTGTGTCAGTGATAA 788
 Qy 1079 AAATGTTTATGATGAAACATACAGATGTTGGGTGGAACGAGATGAGCAAAATATCATGA 1138
 Db 789 GAGCATCTAGACACATACATGTTTTCATGTTTGAAGAGATGAGGAATTTATGCTGA 848
 Qy 1139 TGTGTAACATGCTTTAGCTTTAGCTTTTGGTATTAAGGATCAATGGTATGAAGTTTCCC 1198
 Db 849 CATAACAAATGCAATGGCTTTTGGCCCTGTTTAAAGATGAATGGTATCATGTTTCCCTC 908
 Qy 1199 AGATCCATTTGGTGAATTTACTTAATGAATTTAGTCTTTGAAGACGAAATATGACGCTTTGA 1258
 Db 909 AGTCGAGTTATCTCTGTTGCTGAAGCTTCCAGTTTCCGTGATCACTTCAAGGATATTT 968
 Qy 1259 AACATATCATGCTGCATATATATATACCAAGAGGATTTA-----TCTTCTGAAAA 1310
 Db 969 AAATGATAGAAATCTCTAATCGAATATATACAGGCTTCAAAAGTCAGTAAATCAGAAAA 1028
 Qy 1311 CAAATCTTGAAGTCAGTCAATTTCTTCAAGAGATATATCCATCAATTTCAACAGGCTT 1370
 Db 1029 TGAATCAATCTTAGATAGCATAGGCTCTTGGTCAGGTAGCTTATTTGAAGGAAAGTGTG 1088
 Qy 1371 TCTAATTAATTCACA-----AAGAGGTGGAAGAAATGCTCTTAAGTTCCC 1414
 Db 1089 CTCATAGGGGTGAAGAAAGCTCCATCTTTGAAGAGATGAAGTATGCTTTAAGTTCC 1148
 Qy 1415 TATCAATACCGGTTTGAAGACGATATAACACTAGACGAAATATACAGCTTTACATGTAGA 1474
 Db 1149 CTCTACACACACTGGACGCTCTAGATCAAGAGGAAATTTGAACGTTTTCATGCAAA 1208
 Qy 1475 CAATACAGAAATTTCTGAAGAACTACATATCACTCATCAATATTTAGTAACTGATTAACCT 1534
 Db 1209 GGAATCCAGATGTTAAGACGGAAT---ACTTGCCTTCCATGCCAATCAAGATATCT 1265
 Qy 1535 AAGTTGGCTGTTGAAGATTTCTACACTGCCAATCTATTTATCTGTTGAAGAAATTAAGG 1594
 Db 1266 AGCTTTGGCTGGTGAAGATTTTCAAGTATCTCAATCTATATACAGATGAATTAATTA 1325
 Qy 1595 TCTTTGAAGAGTGGGTGATAGAAATAGTTGGACCACTCAAGTTGCTAGGCAAAAGAC 1654
 Db 1326 TCTTGAAGTGGGTGAAGATGAAGATGAAGAGCTCGATCAGCTGCCATTTTGCACGCCAGAGTT 1385
 Qy 1655 CGCCTACTGTTATTTCTGTTGCTGCAACACTTCTGCTCCCGAATATCAGATGGCG 1714
 Db 1386 GACATATGCTATCTTCTGCTGCTGCTACCATTTCCCGGCGAATGCTGAGGCCG 1445
 Qy 1715 TATTTATGGCGCAAAATGCAATTAATTAACCTACAGTATGTTGATGACTTTTGTATPCGG 1774
 Db 1446 CATTTGATGGCTTAAATGTTGATGACAACTGTTGTTGATGACTTCTTTCATCTGG 1505
 Qy 1775 TGTGTAATCGATGAATTTGACCACTGATTTCAATGTTGTTGAAGAAATGGAATGATAGT 1834
 Db 1506 GGGATCAAGAAAGAACTTGAAGAACCTCTTGGCTTTAGTTTGAAGAGGGGATGACATCC 1565

Qy 1835 CGACAGGATGTTGTTTCAGAGCATGTTCCGATTTTATTTTACATTAAGATGCAT 1894
 Db 1566 AG--AGAGTCTTACTACAGAACCAAGAAAAAGTTTTTTCTGCTATTTATCTACAGG 1622
 Qy 1895 CTGTTGATTTGAGATGAGCTTTTAAATGCAAGCGCGCATCTACTAGCCATGTTAT 1954
 Db 1623 GACCCAGCTTGGAGCAAAAGGCTTCTGCAATTAACAAGCGCGGATGTTTCAAAACCGCTAAC 1682
 Qy 1955 TCAACTTGGTTGGAACCTAATGAATAGTATGTTGAGAGAAGCTATATGCAACAAGATGC 2014
 Db 1683 AGAATATGGTTATGTTCTGATGAGGTCTATGATGACCAAGCTGAATGCGAGAGACAAA 1742
 Qy 2015 TTATGTGCCAACATTAATGAATATATGCAAAACGCTTACGTTGTCATTTGATTTAGGCC 2074
 Db 1743 ATATGTGCGCAATGGAAGAAATATATGCAAAATGCTGTTGCTCATTTGCACTGGGACC 1802
 Qy 2075 GATTGTCAGCCGCTTATTTACTTGTGCGGCCCAATTAATACAGAGGAGATTTGTTGAAAG 2134
 Db 1803 TATTGTGCTCCGACTCTGTTATTTGCTAGGACCGAAGCTCCAGAGGATGTCGTCAGGA 1862
 Qy 2135 CTCTGAATATCATATCTATTTAAGCTAATGAGCACGAGGCTCGACTTTCTAAACGATAT 2194
 Db 1863 TCATGATGATCAATGAATTTATTTAGACTGATGAGCACTTGTGGCGTCTCCTGAATGACAG 1922
 Qy 2195 CCATAGCTTTCAAGAGGGAATTTAAGGAAGGCAAAATTAACCGGCTAGCATTTGATTTGAG 2254
 Db 1923 CCAAGCTTTGAGAGGAGAGGCTTGAGGGAAGCTGAAACAGTCTCTCACTGTTGTTCA 1982
 Qy 2255 TAAACGAGAAAGTGGGAAAGTGGAGAGAGAGGTTGTTGGAGGAGATGATGATGATTA 2314
 Db 1983 TCACAGTGTGTTCTTATCTCCATAGACGAGCTAAATGAAGCCCAAGAAATCCATAGA 2042
 Qy 2315 AAACAGAGGAAAGAAATTAATGAAATTAATTTTGAAGAAATTTGTTAGCATTTGTTCTTAG 2374
 Db 2043 CACTTCCAGAGAACTTGTTAAGATTGTTCTTGGAGAAAGGCT---GCTGTTCTTAG 2099
 Qy 2375 AGCTTGTAAAGATGCAATTTTGAACATGTTGCAAGTGTCAAGTTTATTTTACCAACGA 2434
 Db 2100 GCCATGCAAGCAGCTGTTCTGGAAGATGTTGCAAGATTTGTTCAATGTTCTACTCTCGGAC 2159
 Qy 2435 TGACGGGTTTA 2445
 Db 2160 TGATGGGTTTA 2170

RESULT 11

AK119442
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) cDNA clone:001-133-C06, full insert sequence.
 AK119442
 ACCESSION
 AK119442.1 GI:37989065
 VERSION
 FLI_CDNA; oligo capping.
 KEYWORDS
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Sukayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.

REFERENCE

1
 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ohtsuki, Y., Tsunoda, Y., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurokawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Narioka, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,

Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Itoh, M., Kagawa, I., Kondo, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12869764

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Ikeda, J., Ikeda, J., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahama, Y., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akai, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yokomizo, S., and Yoshimura, A.
Collection, mapping, and annotation of 28K full-length cDNA clones from japonica rice
Unpublished
3 (bases 1 to 3234)
Kikuchi, S.
Direct Submission
Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: shikuchien@affrc.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)
This clone is one of the 32K full-length cDNA clones from japonica rice.
URL : <http://cdna01.dna.affrc.go.jp/cdna/>
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Yamamoto, M., and Nakahama, Y.
FAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Iehibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Oeato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akai, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.
Location/Qualifiers
1. 3234

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-133-C06"

ORIGIN

Query Match 19.0%; Score 530.8; DB 8; Length 3234;
Best Local Similarity 56.8%; Pred. No. 7.6e-94;
Matches 1117; Conservative 0; Mismatches 807; Indels 42; Gaps 6;
QY 513 AAAGATTCTCTATCTTCAACCTTAGCATGTTATTTGATGCTTGTCTTTCAGCTACTGAA 632
Db 957 AGAGATGTTCTGTCTGTCTACGTGTGGCGTGTCTTTCGCACTCAAGAGATGGAATTTGGG 1016
QY 573 GAAGATCAAAATAAATAAGGCTTAAAGTTTATTTAGTCAAAATCTTGTCTTTCAGCTACTGAA 632
Db 1017 CAGGACACATCAGGAGAGGACTAGATTTTATTTGGAAGAAATTTCTCCATTGCCATGGAT 1076
QY 633 AAAAGTCAACCATCTCCATTGGTTTTCACATCATATTTCTCTGTGGTTTCTTGTGATGCG 692
Db 1077 GAGCAGATTGTCTCTCTGTCTTTCACATCACTTTCTCTGTGTATGCTTAGCCTTGCC 1136
QY 693 AAAAATTGGACATAAACCCTCTTTCARAAACAAACAGATTTTATTTGATGCTACATAG 752
Db 1137 ATGGGATGGATTTAGAAATCCCTGTCTCAGACAGACGGATTTGATAGACTTTCTTCCACCTC 1196
QY 753 AGGGAATTGGACCAAAAAA-----GATGCCATTCAAATGAGATGATGATGATGATG 803
Db 1197 CGGAGATAGAACTGGAAAGAGAGGCTGGAGACCATCTTTATGGAAGAAAGCATATATG 1256
QY 804 CGGTATATCTTGAAGGACTCGGTAATTTATGATTGGAATATGTTGGAAGAAATATCAG 863
Db 1257 GCTTATGTCACAGAAGGATTAGGAAACCTGTGTGGATGGGATGAGATTATGATGTTCCAG 1316
QY 864 ATGAAAAATGCTTCTGTTTTCACCTCCACATCAGCAACAGCTGCTGCTTCAATTAATCAT 923
Db 1317 AGHAGAAATGATCATTTTCACTGCTTCCCAACTGCTGCCACATTAGTCAACAC 1376
QY 924 CAAATCTGCTGTTCTTAAATTTAAATTTTAAATTTTGGACAAAGTTTGGTAATGCACTG 983
Db 1377 TATAACGATAAAGCCCTCCAGTACTTAAATTTGCTGTGTAGTAATTTGGCAGTGCAGTA 1436
QY 984 CCAACGTTATCTCTCANGATTTATTTATCCGACTTTTCTATGTTGACCAATTTGAAGA 1043
Db 1437 CCAACGTTATCTCANGATTTATTTATCCGACTTTTCTATGTTGACCAATTTGAAGA 1496
QY 1044 TTAGGAATTTTCAACCATTTTCAAGTGGAAATTTAAATAATGTTTATGATGAAACATACAGA 1103
Db 1497 ATGGGAATTTTCAAGTAAATTTTGTGAGTAAAGGATGATGATGATGATGATGATGATG 1556
QY 1104 TGTGGTGAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1163
Db 1557 TCAATGTTGAAGAGATGAGAAATTTCTGCGACATACGATGATGATGATGATGATGATG 1616
QY 1164 CGGTATTAAGGATCAATGGGTATGAAGTTTCCCGATCCATTTGGCTGAAATTTACTAAT 1223
Db 1617 CGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1676
QY 1224 GAATTAGTTTGAAGAGCAATATGAGCTTCTTGAACATATATGCGTGCATATATTA 1283
Db 1677 GCTTCCAGTTTCCGTTGAATCACTTCAAGGATATTTAAATGATAAGAAATCTTAAATCGAA 1736
QY 1284 TACCAAGAGATTTA-----TCTTCTGGAACAAATCTTGAAGTCAAGTCAAGTATTC 1335
Db 1737 TTATCAAGGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1396
QY 1336 TCAAGAGATTAATTCATCTGATTTCAACAGGCTTTCTAAAT-----TA 1379
Db 1797 TCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1856
QY 1380 ATTCAAAAGAGTGGAAATGCTCTTAAGTTCCCTTATCAATACCGGTTTAGAACCATTA 1439

FEATURES

source

Db 1957 ATCTTTGAAGAGATGAAGTATGCTCTTAAGTTTCCCTTCTACACACACTGGACGCTTA 1916
Qy 1440 AACACTAGAGAAATATACAGCTTTTCAATATGACAAATACAGAAATCTGAAACTACA 1499
Db 1917 GATCACAGAGGAAACATTTGAACGTTTGTATGCAAGGATTCAGAGATTTAAAGACGAA 1976
Qy 1500 TATCACTCATCAATATATAGTAACTGATTAACCTTAAGTTGGCTGTTTGAAGATTTCTAC 1559
Db 1977 T---ACTTGTCTCTCATGCCAATCAAGATATCTAGCTTGGCTGTTTGAAGATTTCACT 2033
Qy 1560 ACCTGCCAATCTATTTATCGTGAAGAAATTAAGAGTCTTGAAGTGGGTGGTGAAGAA 1619
Db 2034 AGTTCTCAATCTATATACAGAGTAACTTAATATCTTGAAGTGGGTGAAGATCAA 2093
Qy 1620 AAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCT 1679
Db 2094 AAGCTCGATCAGCTGCCATTTGCACGCCAGAGTTGACATATGCTACTTTCTGCTGCT 2153
Qy 1680 GCAACACTTTCTGCTCCCGAATATACAGATGCGCTATTTTCATGGCCCAAAAATGGCATA 1739
Db 2154 GCTACCATATTTCCCGTGAATTTGCTGAAGCCGCATTCGATGGCTTAAATATGCTGTA 2213
Qy 1740 TTAACACAGATGTTGATGACTTTTGTATATCGGTGTTACATCGATGAATGACCAAC 1799
Db 2214 CTGACAACTGTTGTTGATGACTTTTGTATCTTGGGGATCAAAAGAGAACTAGAAAAC 2273
Qy 1800 CTGATCAATGTTGTTGAAATGGAATGTAGATGTGCAAGAGATGTTGTTTCAGAGCAT 1859
Db 2274 CTCATTGCTTTAGTTTGAAGTGGGATGGACAT---CAAGAGAGTTCTACTCGAACA 2330
Qy 1860 GTTCGGATTTATTTTATAGCATTAAGATGCAATCTGTTGATTTGGAGATCAAGCTTTT 1919
Db 2331 GTAAAGATGTTTTTTTCTGCTATTTATCTACAGTGAACACCTTGGACCAAGCTTCT 2390
Qy 1920 AATGGCAAGCGCGGATGTAATGACCACTGTTATTTCAAACTGTTGTTGGAATCAATGAT 1979
Db 2391 GCAATTACAGGCGGTGATGTTACAAAACCTTAACAGAAATATGTTATGCTGATGAGG 2450
Qy 1980 AGTATGTTGAGAGAGCTATATGGAACAAGATGCTTATGTCGCAACATTAATGAATAT 2039
Db 2451 TCTATGATCAGCGAAGCTGAATGGCAGAGGACAAAATATGTGCCGCAATGGAAGAAAT 2510
Qy 2040 ATGGAAACGCTTACTGTCATTTGCAATTAGGCGCGATTTGCAAGCGGCTATTTACTTT 2099
Db 2511 ATGGCAAAATGCTGTTGCTCATTTGCTGCTGGACCTATTTGCTCCGACTCTGTATTC 2570
Qy 2100 GTGGGCGCCAAATATCAGAGGAGATTTGTAAGCTCTGAATATCAATATCTATTTAAG 2159
Db 2571 GTAGGACCGAAGCTCCAAGAGGATGTCGTCAAGGATCATGATCAATGAATTTATTA 2630
Qy 2160 CTAATGAGCAGCGAGGTGCACTTTCTAAAGATATCCATAGCTTCAAGAGGGAATTTAAG 2219
Db 2631 CTGATGAGCACTTTGTGGCGCTCTCCTGAATGACAGCAAGGCTTTGAGAGGGAGAGCC 2690
Qy 2220 GAAGGCAAAATTAACCGCGTAGCATTTGATGTAACGGAATAAGTGGGAAGTGGAA 2279
Db 2691 GAGGAAAGCTGAACAGTGTCTCACTGCTTGTTCATCAGTGTGTTCTATCTCATA 2750
Qy 2280 GAAGAGGTTGTGGAGGAGATGATGATGATTAATAAACAAGAGGAAGAAATTAATGA 2339
Db 2751 GACGAGGCTAAATGAATGAAGCCGAAATCCATAGACACTTCCAGGAGGAACTTTGTTA 2810
Qy 2340 TTAATTTTGAAGAAATGTAGCATTTGTTCTTAGAGCTTGTAAAGATGATTTTGAAC 2399
Db 2811 TTGGTCTCTGGAAACAGGGT---GCTGTTCTTAGGCCATGCAAGCAGCTGTTCTGGAAG 2867
Qy 2400 ATGTGTCACGCTGTTGAATTTTTTTTACGCAAGCATGACGGGTTTA 2445
Db 2868 ATGTGCAAGATTTGTTCAATGTTCTACTCTCGGACTGATGGTTTA 2913

RESULT 12
AY347880

LOCUS
DEFINITION
AY347880
2529 bp mRNA linear PLN 05-AUG-2004
Oryza sativa (japonica cultivar-group) map BAC OSUNB00137; BAC
OSUNBA0052P16 ent-kaurene synthase like-4 (OSKS4) mRNA, complete
cds.
AY347880
AY347880.1 GI:34452221
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (Bases 1 to 2529)
Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,E.S. and
Upadhyaya,N.M.
Isolation and characterization of a Ds-tagged rice (Oryza sativa
L.) GA-responsive dwarf mutant defective in an early step of the
gibberellin biosynthesis pathway
Unpublished
2 (Bases 1 to 2529)
Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,E.S. and
Upadhyaya,N.M.
Direct Submission
Submitted (21-JUL-2003) Rice Functional Genomics Group, Genomics
and Plant Development Program, CSIRO Plant Industry, Cnr. Barry
Drive and Clunies Ross Street, GPO Box 1600, Canberra, ACT 2601,
Australia

FEATURES
source

1..2529 Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="4"
/map="BAC OSUNB00137; BAC OSUNBA0052P16"
1..2529
/gene="OSKS4"
1..2529
/gene="OSKS4"
/note="has an extra exon compared to OSUNB0052P16.10 in
GenBank Accession Number AL62936.2; differs from OSKS4
sequence in GenBank Accession Number AY347876"
/codon_start=1
/product="ent-kaurene synthase like-4"
/protein_id="AAQ72563.1"
/db_xref="GI:34452222"

gene

CDS

translation="WASPEAVARSLVAPRRRALGLPAAAAAPFVLCRRHN
GMRPHYSFACSAELDTGEROLPSTGTAVMSSCPGVYEGRMVGNSTQINNGWEAR
IURHLENPEFLFSSTDIAWAVPLPGTDHLQAPCEPCVEWMLQNHSGSWGNEF
DSSAKDILLTLACTIALLEKNVSGEQIRRGHFIATKNFSIVDDQIAAPIGNLTFF
PAMVNAIKMGLFEFFASEISIDQILHDMELKRIAGDESLGREATFYAIEGLEESM
VDMVKPQKNGSLFSPAATAALVHYDDKALGYLSVNVKFGGVPVYPLNI
FSQLSMDPLVNIIGISRHSFSDIKRILDKTYILWSORDEVMIDLPTCAMAFELRMN
GYGSSDDLSHVAEASTPHNSVGVLDLDTLLELYKASKVLSNEPLETKMGKWSG
SLHKEKLCDDIRGTFPILEVEYALKFPFYATLEPLDHKNLENFARAYOKIKTKNM
PCHVEDLLALAAEDFSCSTYQNEISQHLESWEKENKLDQLEFTRKLNLSVLSAA
TTPYELSDARTACASIALTVADDFFDVSGSKSEQENLISLVEKWDQTHKVEFISE
NVKAFYFALSTVQNGAMASAVQRNDVTKNVSWLDYLSRLATDAEQRQSPITM
REYKNSIVTFALGPTILIALYFMQNLWEDIYKNAEYDELFLMNTCGRQLNDIQSF
RECKDKGLNSVLLVLDKSDVMSVEAEKATNESISSORRELLRLVREDGVIKPSK
KEMFNLYKTSIHVYSQADGFSFKEMNGMNGVIFPEPLKTRGN"

ORIGIN

Query Match 17.2%; Score 479.6; DB 8; Length 2529;
Best Local Similarity 53.8%; Pred. No. 8.7e-84;
Matches 1185; Conservative 0; Mismatches 959; Indels 60; Gaps 7;
Qy 285 ACTAATCTGATCATTTGATACCAACAAAGAGATCCAAACACAGTTAAATATGAGAA 344
Db 274 ACAAGTCAGATAAACATGGATGGGAGCTAGATAATCTTAGGCACCTTGGAGAACCCCGAG 333
Qy 345 ATTCTCTGTTCTTCATATGACACAGATGGGTAGCCATGTCCTCTTCACAACTCA--- 401

| | | | | | |
|---|---|--|------|--|------|
| D | b | | 334 | TTCCTTACCAACTCTTTCAATAGCATAGCATGGGTGGCTATGTTGGCCATTGGTGCCAATTGGCGGCACATGAT | 393 |
| Q | y | | 402 | ---CCCAAATCGCCTTGTGTTTCCCTGAGTGTCTCAAATGGTTAAATTAATAATCAATCAAGCTTAAT | 458 |
| D | b | | 394 | CATCTTCAAGCTCCAATGCTTCCCTGAATGTGGAAATGGAATGCTACAACCAACACAGT | 453 |
| Q | y | | 459 | GATGGTTCAATGGGTCTTGTTAATCACTCATAAATCATACACCCGTTGCTTAAAGAT | 518 |
| D | b | | 454 | AATGGGTCTGGGGTGTC-----ATGAATTTGACTCATCAGCAGCAAGGAT | 501 |
| Q | y | | 519 | TCTCTATCTTCAACATTAGCATGTAATTGTGCATTAATAAAGAATGGAATCTTGGGGAAGAT | 578 |
| D | b | | 502 | ATTCTCCTATCACCTTTGCCATGTAATTCACACTTGAGAAATGGGAATCTCGTTCCGGAG | 561 |
| Q | y | | 579 | CAAAATAAATAAAGGCTTAAGTTTTAATGAGTCAAAATCTTGTCTCAGCTACTCAAAAAAGT | 638 |
| D | b | | 562 | CAAAATAAGAGAGAGGATTAATTTTTATCGCAAGAATTTCTCCATTTGTTATGATGACCAG | 621 |
| Q | y | | 639 | CAACCATCTCCCAATTTGGTTTGGACATCATATTTCTGTGTTTGGTTGAGTATCGCAAAAAC | 698 |
| D | b | | 622 | ATTGCTGCACCTATAGGCTTCAACCTCAATTCCTGCTATGTTAAACCTTSCCATTAAG | 681 |
| Q | y | | 699 | TTGGACATAAAACCTCCTCTTTCAAAAACAAACAGAAATTTAGTTTCATGCTCATATAAGAGGAA | 758 |
| D | b | | 682 | ATGGGTTTGGAAATTTCTGCCAGTGAATTAATGATTTGATCAGATTTCCACCTCCGTGAT | 741 |
| Q | y | | 759 | TTGGAGCAAAAAAGA-----TGCCATTCAAATGAGATGGATGGATACATTGGCGTAT | 809 |
| D | b | | 742 | ATGGAAATTGAAAGACTGGCTGTGACGAAATCTTTGGGGAAGAGGCATATTTCCGCTAT | 801 |
| Q | y | | 810 | ATCTCTGAAGACT---CGGTAATTTATATGATTTGGAAATATGGTGAAGAAATATCAGATG | 866 |
| D | b | | 802 | ATTGCTGAAGGCTTAGAAGAAGCATGTTGGATTGGAGTGAAGTTATGAAGTTCACGGG | 861 |
| Q | y | | 867 | AAAAATGTTCTGTTTTCAACTCACCATCAGCAACAGCTGCTGCTTTTCATTAATCATCAA | 926 |
| D | b | | 862 | AAGAATGGAATCATGTTTCACTCCC GGCTGCACTGCTGTGCATTAAGTCCACAGATAC | 921 |
| Q | y | | 927 | AATCCTGGTTGCTTAATTAATTTAAATTCACCTTTTGGACAAGTTTGGTAATGCAGTCCCA | 986 |
| D | b | | 922 | GATGATAAAGCCCTGGGATACCTATATTTCTGTTGTCATAAATTTGGAGGTGAAGTACCA | 981 |
| Q | y | | 987 | ACAGTTTATCTCATGATTTATTTATCCGACTTTCATGTTGATGTCACAAATTTGAAAGATTA | 1046 |
| D | b | | 982 | ACCGTGATATCCGCTAAATATATTTTCTCAGCTTTCATATGGTGSATCTCTCGTCAATTT | 1041 |
| Q | y | | 1047 | GGAATTTCACACCAATTTCAGAGTGGAAATTAATAAATGTTTTAGATGAACACATACAGATG | 1106 |
| D | b | | 1042 | GGAATATCTCGGACTTTCTAGTGAATAAAGCGCATTTTGGATAAGACATACATTTTA | 1101 |
| Q | y | | 1107 | TGGGTGGAACGAGATGAGCAAAATATCATGGATGTTGTAAATGTCGTTTAGCCTTTCCG | 1166 |
| D | b | | 1102 | TGGTCACAGAGAGATGAGGAAGTAAATGCTGGATTTTACCAATGCGCAATGCGATTTCCG | 1161 |
| Q | y | | 1167 | TTATTAGGATCAATGGGTATGAAGTTTCCCAAGATCCATTGGCTGAAATTTACTTAATGAA | 1226 |
| D | b | | 1162 | CTTTTGGGTATGAACGAGATATGGTGTTCCTCAGATCACTTGTCCCATGTGCTGGAAGCC | 1221 |
| Q | y | | 1227 | TTAGCTTTGAAAGCAATATGACG-----CTCTTTGAACA | 1262 |
| D | b | | 1222 | TCAACTTTCATAACTCAGTTTGAAGGATATTTAGATGATACAAATCCTTATTATGAATG | 1281 |
| Q | y | | 1263 | TATCATGGCTCACATATATTATACCAAGAGGATTTATCTTTCTGCAAAACCAATCTTGAAG | 1322 |
| D | b | | 1282 | TACAAAGCTTCAAAAGTCAGTTTATCAGAAAATGAGCCAATCCTAGAGAAATGGGTTC | 1341 |
| Q | y | | 1323 | TCAGCTGATTTCTCAAGAGATTAATATCCACTGATTTCAAACAGGCTTTCTTAAATTAATT | 1382 |
| D | b | | 1342 | TGGTCAGGTAGCTTTATTGAAAGAAAAAATTTGTGCTCCGATGACATCCGAGGAAACCAATC | 1401 |
| Q | y | | 1383 | CACAAAGAGGTGGAATATGCTCTTAAAGTTCCTTATCAATACCGGTTTAGAACGATHAAC | 1442 |
| D | b | | 1402 | CTTCGAGAGTGAGTATATGCTCTCAAAATTTCCAAATTTTATGCCAGCTGGBACCTCTAGAC | 1461 |

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kushimoto, N., Iizaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda R., Yabuchi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Ohneda R.

| | | | |
|----|-----|---|-----|
| QY | 285 | ACTAATCTGATCATTTGATACAACCAAGAACGGATCCAAAAACAGTTTAAAAATGTGAA | 344 |
| Db | 333 | ACAAGTCAGATAAACAATGGGATGGGAGGCTAGAAATACTTTAGGCNCTTGGAGAACCCGGAG | 392 |
| QY | 345 | ATTTCCTGTTCTTCATATGACACAGCATGGGTAGCCATGGTCCCTTCTCCAAATCA--- | 401 |
| Db | 393 | TTCTTACCATCTTCATATGACATAGCANTGGTGGCTATGTTGCCATTCGCCGGCACATGAT | 452 |
| QY | 402 | ---CCCAAAATCGCCTGTTTTCCTGAGTGTCTCAATTGGTTAAATTAAATCAGCTTAAT | 458 |
| Db | 453 | CATCTTCAAGCTCCATGCTTCCCTGAATGTGTGAATGGATCTACAAAAACCAACACAGT | 512 |
| QY | 459 | GATGGTTCAATGGGGTCTTTGTTAATCAACCTCATATATCATATACCCCGTTCCTTAAAGAT | 518 |
| Db | 513 | AATGGGTCGTGGGGTGTCA-----ATGAATTTGACTCATCAGCCAGCAGAGAT | 560 |
| QY | 519 | TCCTCATCTTCAACATTAGCATGATTTGTCATATAAAACATCGAATGTTTGGGGAAGAT | 578 |
| Db | 561 | ATTCTCTATCCACTTTGGCANGATTATTTCACCTTGGAATGGAATGTCTGGTTCGGAG | 620 |
| QY | 579 | CAATAATAAAGGCTAAAGTTTATTTCAGTCAAACTTTGCTCAGCTACTGAAAAAAGT | 638 |
| Db | 621 | CAATAAGGAGAGGATTACATTTTATCGAAAGAATTTCTCCATTGTTATTGATGACCAG | 680 |
| QY | 639 | CAACCATCTCCCATTTGGTTTGGACATCATATTTCTCTGGTTTGGCTTGAGTATGCGAAAAAC | 698 |
| Db | 681 | ATTGCTGCACCTATAGGCTTCAACCTCACATTCCTGCTATAGTTTAACTTGCCCATTAAG | 740 |
| QY | 699 | TTGACATATAACCTCTTTCAAAACAACAGATTTTAGTTTGGTGTACTACATAAGAGGAA | 758 |
| Db | 741 | ATGGGTTTGGAAATTTCTCTGCCAGTGAATTAGTATTGATCAGATTCCTTCACCTCGTGAT | 800 |
| QY | 759 | TTGGNGCAAAAAAGA-----TGCCATTCAAAATGAGATGGATGGATACTTTGCGGTAT | 809 |
| Db | 801 | ATGGAAATTGAAAGAAGCTGGCTGGTGACGAAATCTTTGGGGAAGAGGCATATTTCCGCTAT | 860 |
| QY | 810 | ATCTCTGAAGGACT---CGGTAAATTTATATGATTGGAATATGGTGAAAGAAATATCAGATG | 866 |

QY 633 AAAAGTCAACCATCTCCCATTTGGTTTGGACATCATATTTCTGCTTGGTTTGGTATGCG 692
Db 192 GAGCAGTTTCATCTTCTCTATAGGTTTCAACTTCACCTTCTCTGGTTTCTTAGCCTCGGC 251
QY 693 AAAAATCTGGACATAAAACCTCTTTTCAAAAACAAACAGATTTTGTGTTGATGCTACATAAG 752
Db 252 ATTGATATGGGTTAGAATTTCTGTGAAGACAAATTTGATGCTGTGGCATTTCTCACCGC 311
QY 753 AGGGAATTTGGAGCAAAAAGATGCC-----ATTCAATGAGATGGATGGATCTTG 803
Db 312 CGGAGATGGAATTGAAAAGGCTGCTGTGGATAGTTCTTTTGGAGAAAAGCATATAG 371
QY 804 GCGTATATCTCTGAAGGACTCGTAATTTATATGATTTGGAATATGCTGCAAGAAAATATCAG 863
Db 372 GCTTTTATCCAGAGGATTCGGAATATGCTGGACTGGATCAAGTTATGAAGTTTCAG 431
QY 864 ATGAAAAATGGTCTTGTGTTTCAACTCACCATCAGCAACAGCTGCTGCTTTTCAATTAATCAT 923
Db 432 AGGAAGATGGATCATTTGTTACGCACTCTTCCACACTGCTGTTGCAATTAATCCACAAA 491
QY 924 CAAAATCTCGTGTGCTTAATTTAATTTCACTTTTGGACAGCTTTGGTAATCGAGTC 983
Db 492 TACAACGACCAACCGCTTCAATACCTAAATTTGCTTCAATGAATTTGGCAGTGCAGTA 551
QY 984 CCAACAGTTTATCCTCATGATTTATTTATCCGACTTTCATGTTTGACACAAATTTGAAAGA 1043
Db 552 CAGCAATGATATCCTTCAAGGGTCAATTTGCTAGCTTTCAATGGTGGACGCGCTTGAAAA 611
QY 1044 TTAGGAATTTCAACCATTTCCAGAGTGGAAATTTAAATGTTTATAGTGAACATACAGA 1103
Db 612 ATGGGAATTTCTCAGCGCTTTGTCAGTGAAATAGAAAGCATCTCTGACATGGCATAACAT 671
QY 1104 TGTGGGTGGACGAGATGAGCAATATTCATGATGATTTGTAACATGCTTTTAGCCTTT 1163
Db 672 TGCTGTTACAGAAATGATGAGGAATCTCATGATGGACATAGCAACATTTGCAATGGCAATTT 731
QY 1164 CGGTTATTAAGGATCAATGGGTATGAAGTTTCCACAGATCCATTTGGCTGCAATTAATAAT 1223
Db 732 CGCCTTTTGGAGATGAATGGTTACGATGTTTCTCAGATGAGCTGTCTCAGCTTGTGGA 791
QY 1224 G-----AATTAGCTTTCAAAAGACGAATATGAGCTCTTGAA 1259
Db 792 GCTTCCACTTTCATGATTCACATACAGGATATTTAAATGATACAAAATCCCTACTTGGAA 851
QY 1260 ACATATCATGCTCACA---TATATATACCAAGAGATTTATCTTCTGGAACAAATC 1316
Db 852 TTGTACAAGACCTCAAAAGTCACTTTATCAGAAACGATCTGATCTTAGATCGCATAGGT 911
QY 1317 TTGAAGTCTAGCTGATTTCTTCAAGAGATAATATCCAC----- 1354
Db 912 TCCTGTTCTGGCAACTTATTTGAGGATAGATGCTGTAGTAAGGTGCAAAAGACTCGA 971
QY 1355 -----TGATTTCAACAGGCTTTCTAAATTAATTTCA----- 1386
Db 972 TTTTGGAGAGATGCTGCAAAACAAATTTAAATTTCTCATTTTCACTTTGGAGGTTTCAGTT 1031
QY 1387 -----AAGAGTGGAAAAATGCTTTAAAGTTTCCCTATCATACCGGT 1427
Db 1032 TATGCTCTGTGTTTGTGTTTTCAGATCGATGATGCTGTTAAATTTTCCCTTGTATTCACA 1091
QY 1428 TTAGAAGCATAAACACTAGACGAAATATACAGCTTTTACAAATGTAGACATAACAAGATT 1487
Db 1092 CTGGAGGCTTAGAACACAGAGAAACATCGAACNTTTGATGCTTGGGTTTCTCTGATG 1151
QY 1488 CTGAAACATCATATCATCTCATCAAAATATTAGTAACTGATTAACCTGAAGTTTGGCTGTT 1547
Db 1152 CT---AACAACAAAATCCTCATCTTTTCTGATCAATCAAGAAATCTTAGCTTTGGCAGTC 1208
QY 1548 GAAGATTTCTACACCTCGCAATCTATTATCTGCTGAAGAAATTTAAAGCTTTGAAAGGTGG 1607
Db 1209 GAAGATTTCTAGTTTCTCTCAACGTTTACCGGATGAATCTCGGATCTTAGATTTGG 1268
QY 1608 GTGCTAGAGAAATAGTTGGACCACTCAAGTTTGTCTAGGCAAAAGACCGCTACTGTTAT 1667

Db 1369 GTGAAGGAGAAAGCTGGACCACTACAAATTTGCTCGGCAGAAACTGACATATTTGCTAT 1328
QY 1668 TTCTCTGTGTGTCGAACACTTTTCTCCGAAATTTATCAGATGCGGTATTTTCATGGGCC 1727
Db 1329 CTGTCTGTCTGTCTACCGTATTTTCTTCTGAATTTGTCGCTGCTGCAATTTTCATGGGCC 1388
QY 1728 AAAAATGCAATTAATACTACAGTAGTTGATGACATTTTGTATATGCGTGGTACAAATCAT 1787
Db 1389 AAAAATGGTCTCTCACAACTGTGTTGATGACTTCTTCTGATGTTGGTGGATCAAAAGAA 1448
QY 1788 GAATTGACCAACCTGATTTCAATGTTTGAATAATGGAATGTAGATGTCGACAAAGATGTT 1847
Db 1449 GAAATAGAAAAACCTGATAGCACTAGTTGAGAAATGGCATGGGCACCATGCGAGTTGAGTTC 1508
QY 1848 TGTTCAGAGCATGTTTCGATTTTATTTTAGCATTTAAAAAGATGCAATCTGTTTGGATTGGA 1907
Db 1509 TATTCGGAACAGGTGGAATAATAGTATTTTCTGCTATTTATACAACTGAAACCATCTTGA 1568
QY 1908 GATGAAGCTTTTAAATGGCAAGCGCGCATGTAACTAGCCATGTTTATTCAAACTTTGGTTG 1967
Db 1569 GCAATGGCTTCTGCAGCACAAAGCGCGTGTATCTTACAAACCACTAGTAGAATAATATGGCTG 1628
QY 1968 GAACTAATGAATAGTATGTTGAGCAAGCTATATGCAAGAGATGCTTATGTCACAACA 2027
Db 1629 GATTTGTTAAGATCTATGATGCTCGAGGCAAGATGGCAGAGATGCCAATATGTACAAACA 1688
QY 2028 TTAATGATGAATATATGAAAAACGCTTTACGTTGTCATTTGCAATTAGGCGCCGATTTGCAAGCG 2087
Db 1689 GTTGAAGAAATACATGACAAATGCTGTGTCATTTGCACTGGSCCAATTTGCTGCCCA 1748
QY 2088 GCTATTTTACTTTTGTGGGCCCCAAATTTATCAGAGGAGATTTGTTGAAGCTCTGAATATCAT 2147
Db 1749 GCATTTGATTTTGTAGGGCAAGAGCTATTAGAGCATGCTGTCAAAGATGAAGAGTAGCAT 1808
QY 2148 AATCTATTTAAGCTAATGAGCAGCAGAGCTGCACTTTTAAACGATATCCATAGCTTTCAAG 2207
Db 1809 AAATATTTTAGCTAGTGAGCACTTTCGGGAGCGCTCTCAATGACTACCCAAAGTTTAGAG 1868
QY 2208 AGGGAATTTAAGGAAGGCAAAATTTAAACGGGTAGCATTTGCATTTTGAAGTAAACGAGAAAGT 2267
Db 1869 AGGAAGGCAACCGAGGGAAGCTGATAGTGTCTTCTACTTGTCTCCACAGTGTGGT 1928
QY 2268 GGGAAAGTGGAAAGAGAGGTTGTGGAGAGATGATGATGATGATTAATAACAGAGAGAA 2327
Db 1929 TCTATGTCATAGAGCCGCTAAAGGCAATGCAAGTCCATAGACGCTGTCTAGGAGA 1988
QY 2328 GAATTAATGAATTAATTTTGAAGAAATGGTAGCATTTGTTCTTAGAGCTTCTGAAGAT 2387
Db 1989 GACTTGTGAAGATTTGTTCTT---CAGGAAGAAAGTGTGTTCTTAGGCCATCGAAGAG 2045
QY 2388 GCATTTTGGAAACATGTCACGTTTGAATTTTTTACGCAACGATGACGGTTTGA 2445
Db 2046 CTCCTCTGGAAGATGTTGAAGATCTTCACTTCTGTTTACTCTCAGAATGATGATTGA 2103

RESULT 15

AK068310 3061 bp mRNA linear PLN 24-JUL-2003
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone: J013149A01, full
DEFINITION insert sequence.
ACCESSION AK068310
VERSION AK068310.1 GI:32978328
KEYWORDS FUJ_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1
REFERENCE
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team,
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Nami, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ootomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saitoh, K., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

Science 301 (5631), 376-379 (2003)

2752273

12869764

2 (bases 1 to 3061)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Nami, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, K., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, T., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.

Direct Submission

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL: <http://cdna01.dna.affrc.go.jp/cdna/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Sato, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Nami, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, Y., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,

Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.

Location/Qualifiers

1..3061

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultiVar="Nipponbare"

/db_xref="taxon:39947"

/clone="J013149A01"

ORIGIN

Query Match

Best Local Similarity 16.1%; Score 448.2; DB 8; Length 3061;

Matches 1156; Conservative 0; Mismatches 963; Indels 60; Gaps 7;

QY 308 CAAAGACGGATCCAAACACAGTTTAAATAATGTAGAAATTTCTGTCTTCATATGACAC 367

Db 414 CGAGCGAGTATAAGGAAGCAGCTCCAGGAGTGAATTAATCAACATCCCATATGACAC 473

QY 368 AGCATGGGTAGCCATGGTCCCTTCTCCAAATCACCCTAAATGGCTTTTCCCTGAGTG 427

Db 474 TGCATGGGTAGCTATGGTGGCCATCCGGGGTTCTGCTCACAATCCAAGCTTCCCTCAATG 533

QY 428 TCTCAATTTGGTTAATTAATATCAGCTTAATGATGGTTCATGGGGTCTTGTTAATCAAC 487

Db 534 CGTGCATGGATCTAGAGATCAATGGGACGATGGATCATGGAGTATTTGACGGGTCCAT 593

QY 488 TCATAATCATATCAATCAATCCCGTTTCTTAAAGATTTCTCTATCTTCAACATTAGCATGTATGT 547

Db 594 ATCCACAGCCAA-----CAAGATGTCTATCATCTAGCTAGCATGTGTCT 641

QY 548 TGCATTTAAAGATGGAATGTTGGGAGAGATCAATAAATAAAGGTCTAAGTTTATGGA 607

Db 642 GGCATCAACAAATGGAATGTTGGTAGGAGCACAATCAGGAGAGGACTAAGTTTCATTGG 701

QY 608 GTCAAAATCTTGTCTCAGCTACTGAAAGAGTCAACCATCTCCCATTTGGTTTGCATCAT 667

Db 702 AAGAAATTTCTCGATTCGGATGATGATCAAGCATGTTGCTCTATAGTTTTCGGCATCAC 761

QY 668 ATTTCTGGTTTGTCTGAGTATCGAAAAAATTTGGACATAAACTCTCTTTCAAAACAAAC 727

Db 762 TTTCCAGCTATGTAACCCCTTGCCATGGTTCGGCTTGGAGTCCCGCTCAGACAAA 821

QY 728 AGATTTTGTGATGCTACATAGAGGGAATTTGGACAAAAAAGATG-----CCA 778

Db 822 CGATTTTGTAGTCTTAATCACTTCGGGAGATGAAAAATACAAAGAGAGCGCGGAACCA 881

QY 779 TTCAAAATGAGATGGATGCTATCTCTGAGGACTCGGTAATTTATATGA 838

Db 882 TTTCTGGGAGAAAGCCCTATATGGCTATCTGGCAGAAAGGATTTGGGAACCTTCTGGA 941

QY 839 TTGGAATATGGTCAAGAAATATCAGATGAAAAATGTTCTGTGTTTCAACTCACCATCAGC 898

Db 942 GTGGGATGAATCATGATGTTTCCAGAGGAAGACGGGTCTATTGTTCACTGTCTTCTTC 1001

QY 899 AACAGCTGCTGCTTCAATTAATCATCAAAATCTGTTGTCTTAATTTAATTAATCACT 958

Db 1002 AACTGCTGGGCAATTAGCCCAATTTACCACGACGATAAAGCTCTCCAATCTTGCATCTCT 1061

QY 959 TTTGGCAAGTTTGGTAATGCTGCTCCCAACAGTTTATCTCATGATTTATTTATCCGACT 1018

Db 1062 AGTCAATAATTTGACGGGTAGTGCACACCTGTATCCATTAACATATTTGTGAGCT 1121

QY 1019 TTCTATGTTGACACAAATTTGAAAGATTAGGAATTTTCCACCACTTTTCCAGAGTGAATAA 1078

Db 1122 TTCAATGTTGGATGCGCTTGAACATCGGAATATCTCAGTATTTTCCAGTGAATAA 1181

QY 1079 AATGTTTGTAGTGAACATACAGATGTTGGTGGAAACGAGATGACCAATATTTATGGA 1138

Db 1182 GAGCATCTTGCATGACATACAGATTTCTCGTTAGGGAAGATGAGGAATCATGTTTGA 1241

QY 1139 TGTGTGAATGCTGCTTTAGCCCTTTCGGTTATTAAGGATCAATGGGTATGAAGTTTCCCC 1198

Db 1242 CGTAAACAGTCGCAATGGCAATTCGCTTTTAAAGATGAACGGTTACGATGTTTCTC 1301
Qy 1199 AGATCCATGTCGTAATAATTACTAATGAATTAAG-----CTTT 1234
Db 1302 AGATGAGCTATCTCATGTTGCTGGAGCTTCAGGCTTCGGTGAATTCGGTTCAAGGGTATCT 1361
Qy 1235 GAAAGACGAATATGAGCTCTTGAACAATATATCGCTC---ACATATATTTATACCAAGA 1291
Db 1362 AAACGATAGAAAATCTGTACTGGAAGTATACAGACCTCGAAACACAGTATATCAGAAA 1421
Qy 1292 GGATTTATCTTTGGAACAAATCTTGAAGTCAAGCTGATTTCTCTCAAGAGATAATATC 1351
Db 1422 TGAATTTGATCTTAGATAGCATAGGCTCTTGTGAGGAGCTATTTGAAGGAAATGCTGTG 1481
Qy 1352 CACTGATTCAAACAGGCTTTCTAAATTAATTCACAAAGGTTGGAATGCTCTTAAGTT 1411
Db 1482 CTCCTAAATGGGAAAGGAATCC-----AGGCGTGAAGAGATCGAGTTTGTCTGAAGTA 1535
Qy 1412 CCCTATCAATACCGGTTTAGAAGCATAAACACTAGACGAATATACAGCTTTTACAATGT 1471
Db 1536 TCCCTTCTACAGCACATTTGGAACGCTAGTCCACAGGAAGACATTTGCTTTTCAGCG 1595
Qy 1472 AGACAAATACAGAAATCTGAAATCTACATPATCACTCATCAATATTTAGTAACACTGATTA 1531
Db 1596 TAAGGGCTCCAGATGCTGAAACACAGATGCAATGCTGTTCAATGACAGTCA---AGATTT 1652
Qy 1532 CTTAAGGTTGGCTGTGAAGATTTCTACCTGCGCAATCTATTTATCTGTGAAGAAATTA 1591
Db 1653 TCTAGCTTTGGCTGTGATGATTTCTGCACTCTCTCAATCTAATTAACAGAAATGAATGAA 1712
Qy 1592 AGGCTTTGAAGGTTGGGTGAGAGATAAGTTGGACAGCTCAAGTTTGTAGGCAAAA 1651
Db 1713 TTATCTCGAAGTTGGGTGAAAGACAAACAGCTCGACAGCTACATTTGCGACGCGAGAA 1772
Qy 1652 GACCGCTACTGTATTTCTCTGTGCTGCAACACTTTTCTGCTCCGAATTAATCAAGTGC 1711
Db 1773 GATAACATATGCTATCTCTCTGCTGCTGCCACCAATTCCTGCTGAAATGGGCTACGC 1832
Qy 1712 GGGTATTTCAAGGCGCAAAAATGGCAATTAATCACTACAGTAGTTGATGACTTTTGTATAT 1771
Db 1833 TCGCACCTCGTGGGCAAGAACCGCTTGGTTGAGGCTGTATTCGACGATCTCTTCGATGT 1892
Qy 1772 CGGTGTGACAACTCGATGATTTGACCAACCTGATTTCAATGTTGAAATTTGAATGTAGA 1831
Db 1893 TGGTGAATTAGAACAAAGAACCAAGAAACCTCTAGCATTTAATGAGAGATGGGAGAGCC 1952
Qy 1832 TGTGGAACAAGGATTTGTTGTTGAGAGATGTTCCGATTTTATTTTATGACATTAAGATGC 1891
Db 1953 TGGTGAAGATGAGTACTACTCTGAAGATGTAAAGATCGTGTCCAGGCTCTTTATAATAC 2012
Qy 1892 AATCTGTTGGATTTGGAGATGAAGCTTTTAAATGGCAAGCGCGGATGTAAGTACCATGT 1951
Db 2013 GGTGAACGAGATTTGGCGCAAGGCTTCCGCGTTTACAAAGCCATGATTTACCAAGTACCT 2072
Qy 1952 TATTTCAAACTTGGTTGGAATTAATAGTATGTTGAGAGAGCTATATGGAACAAGAGA 2011
Db 2073 AGTAGAGTATGCTACATGTTGTTGAGGTTGATGAAGTTGAGGCAATGGCAGAGAG 2132
Qy 2012 TGCTTATGTGCCAACATTAATGAATATATGAAAGCGTTTACGTGTCATTTGCAATAGG 2071
Db 2133 CCAGCATCTGCCAAAGTTTGAAGAAATACATGAAAGTGGCATGTTGCTCACTTTGACAGAGG 2192
Qy 2072 CCGATTTGCAAGCGGCTATTTACTTTTGTGGGGCCAAATATATCAGAGGATTTGTTGA 2131
Db 2193 CTGTACGGTGTATGTCGGCATTTGTTCTGTATCGGAGAAAGCTCCCGAGGGGTATTTGCGA 2252
Qy 2132 AAGCTCTGAATATCATTAATCTATTTAAGCTTAATGAGCACGAGGCTCGACTTCTTAAACGA 2191
Db 2253 ACTTGAAGAAATACGATGAGTTGTTTCAGCTGATGGGCACCTTGGCGCGCTCTCTTGAATGA 2312
Qy 2192 TATCCATAGCTTCAAGAGGGAATTTAAGGAGGCAAAATTAACCGCGTAGCATTTGCAATTT 2251
Db 2313 CATTCGAGGCATTCAGAGGGAGGAGTCGGACGGCAAAATGACGAACGGCGTCTCGCTGCT 2372

Qy 2252 GAGTAAACGAGAAAGTGGGAAAGTG---GAAGAAGAGGTTGAGGAGATGATGATCAT 2308
Db 2373 CGTTCA CGCTAGCGCGCTCCATGTCGTTAGACGAGGCCAAAACCGAAGTGTATGAAGCG 2432
Qy 2309 GATTAAAAAACAAGAGGAAAGAAATTAATGAATTAATTTTGAAGAAATGCTAGCATTTGT 2368
Db 2433 CATCGACGCTTAACGAGGAAAGCTGCTGAGCTTGGTTCGTCAGCGAACAGGAGGCCCTAT 2492
Qy 2369 TCCTAGAGCTTGTAAAGATGCAATTTTGGAAACATGTCACGTTTGAATTTTATTTTACGC 2428
Db 2493 CCCGAGGCGGTGCAAGCAGCTGTTCTGGAAGATGTGCAAGATTTCTTACCTGTTCTACTA 2552
Qy 2429 AAAAGATGACGGGTTTACT 2447
Db 2553 CCAGACCGAGCGATTCAGT 2571

Search completed: October 16, 2004, 23:04:19
Job time : 11760.6 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:30:51 ; Search time 124.839 Seconds
(without alignments)
1543.991 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINNDPVWSSQNE.....DTATNLHDELIIYIDHSEL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|------------|--------------------|
| 1 | 1720 | 100.0 | 335 | 2 | Q12051 | Q12051 saccharomyc |
| 2 | 838 | 48.7 | 332 | 2 | Q6FRW9 | Q6frw9 candida gla |
| 3 | 753 | 43.8 | 320 | 2 | Q758X0 | Q758x0 ashbya goss |
| 4 | 753 | 43.8 | 320 | 2 | RA552447 | RA552447 ashbya go |
| 5 | 634 | 36.9 | 396 | 2 | Q87411 | Q87411 emericeia |
| 6 | 629 | 36.6 | 327 | 2 | Q6C8T9 | Q6c8t9 yarrowia li |
| 7 | 628.5 | 36.5 | 418 | 1 | Q6C8T9 | Q6c8t9 yarrowia li |
| 8 | 625.5 | 36.4 | 375 | 2 | Q9C452 | Q9c452 penicillium |
| 9 | 623 | 36.2 | 428 | 1 | Q6PP NEUCR | Q6pp neucr |
| 10 | 614 | 35.7 | 387 | 2 | Q6MVH1 | Q6mvh1 aspergillus |
| 11 | 614 | 35.7 | 387 | 2 | CAF32032 | CAF32032 aspergill |
| 12 | 612.5 | 35.6 | 343 | 2 | Q6F6E2 | Q6f6e2 phoma betae |
| 13 | 598.5 | 34.8 | 343 | 2 | Q6F5B6 | Q6f5e6 phoma betae |
| 14 | 591.5 | 34.4 | 303 | 2 | Q9P885 | Q9p885 mucor circi |
| 15 | 583 | 33.9 | 300 | 1 | GGPP HUMAN | GGpp h human |
| 16 | 581 | 33.8 | 300 | 2 | Q6NM19 | Q6nm19 homo sapien |
| 17 | 581 | 33.8 | 300 | 2 | RAH67768 | RAh67768 homo sapi |
| 18 | 578 | 33.6 | 294 | 1 | GGPP BOVIN | GGpp bovin |
| 19 | 577 | 33.5 | 300 | 1 | GGPP MOUSE | GGpp mouse |
| 20 | 577 | 33.5 | 300 | 2 | AAH69913 | AAh69913 mus muscu |
| 21 | 576 | 33.5 | 315 | 2 | Q6CM36 | Q6cm36 kluyveromyc |
| 22 | 570 | 33.1 | 300 | 2 | Q6F596 | Q6f596 rattus norv |
| 23 | 558.5 | 32.5 | 327 | 2 | Q7ZYH0 | Q7zyh0 brachydanio |
| 24 | 554 | 32.2 | 392 | 2 | Q9P965 | Q9p965 nigrospora |
| 25 | 553 | 32.2 | 371 | 2 | Q9C446 | Q9c446 penicillium |
| 26 | 545.5 | 31.7 | 338 | 2 | Q61539 | Q61539 drosophila |
| 27 | 545.5 | 31.7 | 338 | 2 | Q9V854 | Q9v854 drosophila |
| 28 | 541.5 | 31.5 | 346 | 2 | Q6BN82 | Q6bn82 debaryomyce |
| 29 | 527 | 30.6 | 335 | 2 | Q86F69 | Q86f69 schistosoma |
| 30 | 524 | 30.5 | 236 | 2 | Q9C226 | Q9c226 mus muscula |
| 31 | 513 | 29.8 | 335 | 2 | Q61538 | Q61538 drosophila |

32 500 29.1 357 2 Q7QI19
33 454.5 26.4 342 2 Q74289
34 337 19.6 347 2 O51294
35 310.5 18.1 327 1 IDSA_METJA
36 282 16.4 323 2 Q8D276
37 282 16.4 324 2 Q7VQN3
38 274.5 16.0 364 2 O05708
39 268 15.6 350 2 Q73MH1
40 268 15.6 350 2 AAS12054
41 262 15.2 317 2 Q6M172
42 262 15.2 317 2 CAF29601
43 249 14.5 742 2 Q9NEB3
44 246.5 14.3 322 2 Q8QI11
45 244 14.2 324 1 IDSA_METTM

ALIGNMENTS

RESULT 1

Q12051
ID Q12051 PRELIMINARY; PRT; 335 AA.
AC Q12051;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Btslp: geranylgeranyl diphosphate synthase (Btslp).
GN Name=BTS1;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Ty4;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,
RA Araujo R., Aparicio A., Barrell B., Badcock K., Benes V., Botstein D.,
RA Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E.,
RA Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S.,
RA Delius H., DiPaolo T., Dubois E., Dusterhoft A., Duncan M., Floeth M.,
RA Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Hebling U.,
RA Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R.,
RA Johnston N., Kaiman S., Klein K., Komp C., Kurdi O., Lashkari D.,
RA Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F.,
RA Mewes H.W., Mirtipati S., Moestl D., Muller-Auer S., Namath A.,
RA Nentwich U., Oefner P., Pearson D., Petel P.X., Pohl T.M.,
RA Purnelle D., Schafer M., Scharfe M., Scherens B., Schramm S.,
RA Schroeder M., Sdicu A.M., Tettelin H., Urrestazu L.A., Ushinsky S.,
RA Vierendeels F., Visers S., Voss H., Walsh S.V., Wambutt R., Wang Y.,
RA Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H.,
RA Hani J.;
RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:0-0(0).
[2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Ty4;
RX MEDLINE=95394944; PubMed=7665600;
RA Winnett E., Ahmed A., Bussey H., Fortin N., Friesen J.D., Hall J.,
RA Storms R.K., Vo D.H., Wang Y.;
RN Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Ty4;
RX MEDLINE=95394944; PubMed=7665600;
RA Jia Y., Cherry J.M.;
RN Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC MEDLINE=95394944; PubMed=7665600;
RX Jiang Y., Proteau P., Poulter D., Ferro-Novick S.;
RA "BTS1 encodes a geranylgeranyl diphosphate synthase in Saccharomyces
cerevisiae.";
RT J. Biol. Chem. 270:21793-21799 (1995).
RL EMBL; U93205; AAB68296.1; -.

ORNames=YALI0D170509;
 Yarrowia lipolytica (Candida lipolytica).
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Dipodascaceae; Yarrowia.
 NCBI_TaxID=4952;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CLIB99;
 RC GENOLEVURES;
 RG Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talia E.,
 RA Goffard N., Frangin L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
 RA Boissame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaut J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Wincker P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CLIB99;
 RC Genoscope;
 RA Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; CR382130; CAG81114.1; --
 SQ SEQUENCE 327 AA; 37316 MW; A982D91858A2F302 CRC64;
 Query Match 36.6%; Score 629; DB 2; Length 327;
 Best Local Similarity 40.8%; Pred. No. 4.8e-36;
 Matches 131; Conservative 62; Mismatches 104; Indels 24; Gaps 6;
 QY 14 WVSSQESLISKYPNHLKPGKFNLIQVIRNVLNPKDQALVSOIVELLHNSLL 73
 DB 11 IWGKAADTALLGFPYLANRGNHREHLIAFGAVIKVDSLETISHITKLHNSLL 70
 QY 74 IDIEDNAPLRGQTTSHLIFGVPTINTANYMYFRAMQVLSQLTKEPLYHNLITFNE 133
 DB 71 VDDVEDNSMLRGLPAHCLFGVPQTINSANYMYFVALQEVKLKS----YDAVSIFTE 125
 QY 134 EILNHRGQGLDIYWRDPEIPIPTQEMLYNMVNTKGLFRLTIRLMEALSPSSHCHS 193
 DB 126 EMINLHRGQGLDIYWRDPEIPIPTQEMLYNMVNTKGLFRLTIRLMEALSPSSHCHS 182
 QY 194 LVPPF-----INLIGLIYQIRDDYLNLDKDFQMSSEKGFADITCKLSFPPIVHALNFTKTK 248
 DB 183 KINFDTLHTLDTLGVYIQLDDYLNLDKDFQMSSEKGFADITCKLSFPPIVHALNFTKTK 237
 QY 249 GQTEQNEILRIILLRTSDKDIKLQILEFPTNSLAYTKNFINOINMKNENKYL 308
 DB 238 -TNPDNHEILNLIKQRTSDASLKKYADVYWRTEKTSFDYCLKRIQAM-----SLKASSYI 291
 QY 309 PDLASHSDPATNLDHLLYII 329
 DB 292 DDLAAAGHDVSKRLAILHYFV 312
 RESULT 7
 GGPP-GIBFU STANDARD; PRT; 418 AA.
 ID GGPP-GIBFU
 AC Q92236;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPSASE)
 DE (Geranylgeranyl diphosphate synthase) [includes:
 DE Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase
 DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
 Name=GGGS; Synonyms=GGPPS;
 Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium moniliforme).
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
 NCBI_TaxID=5127;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=m567;
 RC MEDLINE=97374453; PubMed=9230902;
 RA Mende K., Honann V., Tudzynski B.;
 RT "The geranylgeranyl diphosphate synthase gene of Gibberella fujikuroi: Isolation and expression.";
 RL Mol. Gen. Genet. 255:96-105(1997).
 CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.
 CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.
 CC -!- PATHWAY: Gibberellins biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
 EMBL; X96943; CAA65644.1; --
 InterPro; IPR000092; Polyprenyl_synt.
 InterPro; IPR008949; Terpenoid_synt.
 Pfam; PF00348; polyprenyl_synt; 1.
 DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 KW Isoprene biosynthesis; Transferase.
 FT ACT_SITE 269 269 By similarity.
 SQ SEQUENCE 418 AA; 46477 MW; BE2A0FEEAD176C9 CRC64;
 Query Match 36.5%; Score 628.5; DB 1; Length 418;
 Best Local Similarity 44.7%; Pred. No. 7e-36;
 Matches 127; Conservative 58; Mismatches 80; Indels 19; Gaps 6;
 QY 15 WSSQESLISKYPNHLKPGKFNLIQVIRNVLNPKDQALVSOIVELLHNSLLI 74
 DB 122 WTDEKENVVRGPDYVISHPGKDFRAQLIGAFNVLDVPTSSLEVIRVVGMLHESLLI 181
 QY 75 DIEDNAPLRGQTTSHLIFGVPTINTANYMYFRAMQVLSQLTKEPLYHNLITFNEE 134
 DB 182 DVQDSSELRGFPVAHIFGVAQTINSNYIYFVALQELHKNLP-----LITIFSDE 236
 QY 135 LILNHRGQGLDIYWRDPEIPIPTQEMLYNMVNTKGLFRLTIRLME--EALSPSSHCH 192
 DB 237 LVNLRHGQGLDIYWRDPEIPIPTQEMLYNMVNTKGLFRLTIRLME--EALSPSSHCH 291
 QY 193 SLVFPNLIIGLIYQIRDDYLNLDKDFQMSSEKGFADITCKLSFPPIVHALNFTKGTQTE 252
 DB 292 DCVPLVNLIGLIYQIRDDYLNLDKDFQMSSEKGFADITCKLSFPPIVHALNFTKGTQTE 349
 QY 253 QHNEILRIILLRTSDKDIKLQILEFPTNSLAYTKNFINOINQIV 296
 DB 350 ----LNLKQKTSDTQIKRYAVAYME-STGSPFYTRKVLVLI 388
 RESULT 8
 Q9C452 PRELIMINARY; PRT; 375 AA.
 ID Q9C452
 AC Q9C452;

01-JUN-2001 (TRENDEL. 17, Created)
 01-JUN-2001 (TRENDEL. 17, Last sequence update)
 01-OCT-2003 (TRENDEL. 25, Last annotation update)
 Geranylgeranyl pyrophosphate synthase.
 Name=ggs1;
 NCBI_TaxID=70109;
 SEQUENCE FROM N.A.
 MEDLINE=21106005; PubMed=11169115;
 Young C., McMillan L., Telfer E., Scott B.;
 "Molecular cloning and genetic analysis of an indole-diterpene gene
 cluster from *Penicillium paxilli*,"
 Mol. Microbiol. 39:754-764(2001).
 ENBL: AF279807; AAK11525.1;
 GO: GO:0008229; P:isoprenoid biosynthesis; IEA.
 InterPro: IPR000092; Polyprenyl_synth.
 InterPro: IPR008949; Terpenoid_synth.
 Pfam: PF00348; polyprenyl_synth. 1.
 PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
 PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
 SEQUENCE 375 AA; 41505 MW; ACA6931809643B05 CRC64;
 Query Match 36.4%; Score 625.5; DB 2; Length 375;
 Best Local Similarity 42.9%; Pred. No. 9.9e-36;
 Matches 126; Conservative 62; Mismatches 89; Indels 17; Gaps 6;
 QY 2 EAKTDELINDPWWSSQNESLISKPNHLLKPKGNFLNLIQVNRVNNLPKDLQAIYS 61
 Db 55 KSSVDGTYKIDGTWSSNKELKIDFYDMHQHQPQKVRQLQIAFNSWQLWPPELSAII 114
 QY 62 QIVELLHNSLLIIDIDNAPLRGQTTSLHFGVPSTINTANYFAMOLVSLTKE 121
 Db 115 KVTMLTASLLIDVDNSVLRGVPVAMNIFTAQTINSANVYFLAQEIQL--KN 172
 QY 122 PLVHNLITPNEELINLHRRGGDIYWRDFELPEIPTQEMYNMKNKGLFLRLTLM 181
 Db 173 PA--AIDIVKELLNLRHGGQDLFWRDL--TCPTDEFLVAVGNKTKGLFLFLAVKLM 227
 QY 182 EALSPSHHGHSLVFFNLGIYQIRDDYLNKDFQMSSEKGAEDITGKLSFPIVHA 241
 Db 228 QA---ESSTGKDCVSLNVNMGFLPQICDDYLNLSNTYTHNKGLEDLTGKFSFPIHS 284
 QY 242 LNFYTKGQEQHEILRLILRTSDKDKLKLQILEFDNSLAYTKNFQNL 295
 Db 285 IR-----SNPGHQLVSLKQKTDSEVRKYAVQVMQ--STGSPTHRQVVRDL 331
 RESULT 9
 GGPP_NEUCR
 ID -GGPP_NEUCR STANDARD; PRT; 428 AA.
 AC P24322; Q7RYC0;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPase)
 DE (Geranylgeranyl diphosphate synthase) (albedo-3 protein) [includes:
 DE Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase
 DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
 GN Name=al-3; ORFNames=B8F8.010, NCU01427.1;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 SEQUENCE FROM N.A.
 STRAIN=74-OR23-1A / FGSC 987;
 RC MEDLINE=91170267; PubMed=1826006;
 RX Carattoli A., Romano N., Ballarín P., Morelli G., Macino G.;
 RT "The Neurospora crassa carotenoid biosynthetic gene (albedo 3) reveals
 RT highly conserved regions among prenilyltransferases."

J. Biol. Chem. 266:5854-5859(1991).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=74-OR23-1A / FGSC 987;
 RC MEDLINE=22542210; PubMed=12655011;
 RX Manhaupt G., Montrone C., Haase D., Meves H.-W., Aign V.,
 RA Hoheisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
 RA Schulte U.;
 RT "What's in the genome of a filamentous fungus? Analysis of the
 RT Neurospora genome sequence.";
 RL Nucleic Acids Res. 31:1944-1954(2003).
 [3]
 SEQUENCE FROM N.A.
 STRAIN=74-OR23-1A / FGSC 987;
 RC PubMed=12712197; DOI=10.1038/nature01554;
 RX Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
 RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
 RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
 RA Qui D., Iankiev P., Bell-Pedersen D., Nelson M.A.,
 RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
 RA Zelter A., Schulte U., Kothe G.O., Jedd G., Meves H.-W., Staben C.,
 RA Marcotte E., Greenberg D., Roy A., Foley K., Maylor J.,
 RA Stange-Thomann N., Barrett R., Guerre S., Kamal M., Kamvysselis M.,
 RA Mauceli E., Bielke C., Rudd S., Friseman D., Krystofova S.,
 RA Rasmussen C., Metzner R.L., Perkins D.D., Kroken S., Cogoni C.,
 RA Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
 RA Desouza C.P., Glass L., Orbach M.J., Berglund J.A., Voelker R.,
 RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
 RA Natvig D.O., Alex L.A., Manhaupt G., Ebbole D.J., Freitag M.,
 RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
 RT "The genome sequence of the filamentous fungus *Neurospora crassa*.";
 RL Nature 422:859-868(2003).
 CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of
 CC IPP onto DMAPP to form geranylgeranyl pyrophosphate.
 CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
 CC = diphosphate + trans-trans-farnesyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranylgeranyl diphosphate + isopentenyl
 CC -!- PATHWAY: Carotenoid biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- INDUCTION: By blue light.
 CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL: U20940; AAC13867.1; --
 CC EMBL: BX294018; CAD70868.1; --
 CC EMBL: ABX01000298; EAA31459.1; --
 CC PIR: S15662; S15662.
 CC InterPro: IPR000092; Polyprenyl_synth.
 CC InterPro: IPR008949; Terpenoid_synth.
 CC Pfam: PF00348; polyprenyl_synth. 1.
 CC PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
 CC PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
 CC Carotenoid biosynthesis; Isoprene biosynthesis; Transferase.
 CC ACT SITE 268 268 By similarity.
 CC SEQUENCE 428 AA; 47887 MW; 7989DERABD8E360F CRC64;
 Query Match 36.2%; Score 623; DB 1; Length 428;
 Best Local Similarity 39.1%; Pred. No. 1.8e-35;
 Matches 125; Conservative 74; Mismatches 99; Indels 22; Gaps 7;
 QY 15 WSSQNESLISKPNHLLKPKGNFLNLIQVNRVNNLPKDLQAIYSQIVELHNSLLI 74
 Db 121 WSEKEKVLITGTYDLYNGHFGKDIQSQWKAFDAMLDIVFSESLEVTIKVISMULTASLLV 180

```

QY 75 DDIEDNAPLRGGTSHLIFGVPSSTINTANYMFRAMOLVSQLTKEPLVHNLITIFNEE 134
Db 181 DVEDNSVLRGFFVAHSIFGIPOTINTSNVYFALQELQKL--KNP--KAVSIIFSEE 235
QY 135 LINLRHGQGLDIYWRDFPEIIPQEMYLNMVNMKGTGLFRLLRLMEALSPSSHGHSL 194
Db 236 LINLRHGQGLDFWRDTL--TCPTEDDYLMVSNKTYGGLFRGLGKLMQAESRSP---VDC 290
QY 195 VPFNLGLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQTOH 254
Db 291 VPLNVLIIGLFOIADDDYHNLNREYTNKMGCEDLTEGKFSFPVHISIR-----SNPSN 344
QY 255 NEILRLILRTSDKDIKLIQLLEFDNLSAYTKNFQNLVNMKNENKYLPLDLASH 314
Db 345 MOLLNLKQKTGDEVEKYAYAYME-STGSFAYTRKVVREF-----DRARQMTEDIDDG 398
QY 315 SDTATNLHDELIIYIDHLS 334
Db 399 RKSGGIHKILDRIMLHQB 418

RESULT 10
Q6MYH1
ID CAF32032 PRELIMINARY; PRT; 387 AA.
AC Q6MYH1.
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Geranylgeranyl pyrophosphate synthetase, putative (EC 2.5.1.1).
GN ORFNames=AFA63.050C;
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
EX PubMed=1498527;
RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,
RA Foster N., Fraser A., Harris D., Larke N., Murphy L., Humphray S.,
RA O'Neil S., Pettea M., Price C., Rabbinswitsch E., Rajadream M-A.,
RA Salzberg S., Saunders D., Seegar K., Sharp S., Warren T.,
RA Denning D.W., Barrell B., Hall N.;
RA "Insight into the genome of Aspergillus fumigatus: analysis of a 922
RT kb region encompassing the nitrate assimilation gene cluster.";
RL Fungal Genet. Biol. 41:443-453(2004).
DR EMBL; BX649606; CAF32032.1; -.
DR GO; GO:0004161; F:dimethylallyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
KW Transferase.
SQ SEQUENCE 387 AA; 43397 MW; 04F3F846COA2A1D7 CRC64;

Query Match 35.7%; Score 614; DB 2; Length 387;
Best Local Similarity 40.8%; Pred. No. 6.6e-35;
Matches 128; Conservative 62; Mismatches 88; Indels 36; Gaps 6;

QY 12 DPWSSONESLSKPYNHLLKPKGNFRNLIVQINRNMVLPKDLQALVSIQVIELLHNS 71
Db 67 DGNWSQVEYKILMGPDYMQHQHKGDIRQLITAFNMVLQVRPESLSITKVVGLHTAS 126
QY 72 LLIDIEDNAPLRGGTSHLIFGVPSSTINTANYMFRAMOLVSQLTKEPLVHNLITIF 131
Db 127 LLVDDVEDNSILRRGIPVAHNIFGTAQTINSANYVFLALQEVQKLNPP-----TAIDIF 181
QY 132 NEELNLHRCQGLDIYWRDFPEIIPQEMYLNMVNMKGTGLFRLLRLMEALSPSSHGH 191
Db 182 VQELLNLHRCQGLDFWRDTL--TCPTEDDYLDWVGNKTYGGLFRLLAQLMQAESIT--G 236
QY 192 HSLVFPFNLGLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQ 251
Db 237 IDCVSIVNMVGLIFQICDDYLNLSNKTYTNQKGLCEDLTEGKFSFPVHISIR-----SN 290
QY 252 EQHNEILRLILRTSDKDIKLIQLLEFDNLSAYTKNFQNLVNMKNENKYLPLDL 311
Db 291 PRNHQISILKORTKDEVEKLYAISYME-STGSFAYTRKVVREF----- 333
QY 312 ASHSDTATNLHDEL 325
Db 334 ---RDKALSLIDEI 344

RESULT 12
Q6F6E2
ID Q6F6E2 PRELIMINARY; PRT; 343 AA.
AC Q6F6E2;

```

```

QY 192 HSLVFPFNLGLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQ 251
Db 237 IDCVSIVNMVGLIFQICDDYLNLSNKTYTNQKGLCEDLTEGKFSFPVHISIR-----SN 290
QY 252 EQHNEILRLILRTSDKDIKLIQLLEFDNLSAYTKNFQNLVNMKNENKYLPLDL 311
Db 291 PRNHQISILKORTKDEVEKLYAISYME-STGSFAYTRKVVREF----- 333
QY 312 ASHSDTATNLHDEL 325
Db 334 ---RDKALSLIDEI 344

RESULT 11
CAF32032
ID CAF32032 PRELIMINARY; PRT; 387 AA.
AC CAF32032;
DT 13-MAY-2004 (TREMBlrel. 27, Created)
DT 13-MAY-2004 (TREMBlrel. 27, Last sequence update)
DT 13-MAY-2004 (TREMBlrel. 27, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase, putative (EC 2.5.1.1).
GN AFA6E3.050C.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,
RA Foster N., Fraser A., Harris D., Larke N., Murphy L., Humphray S.,
RA O'Neil S., Pettea M., Price C., Rabbinswitsch E., Rajadream M-A.,
RA Salzberg S., Saunders D., Seegar K., Sharp S., Warren T.,
RA Denning D.W., Barrell B., Hall N.;
RA "Insight into the genome of Aspergillus fumigatus: analysis of a 922
RT kb region encompassing the nitrate assimilation gene cluster.";
RL Fungal Genet. Biol. 41:443-453(2004).
DR EMBL; BX649606; CAF32032.1; -.
KW Transferase.
SQ SEQUENCE 387 AA; 43397 MW; 04F3F846COA2A1D7 CRC64;

Query Match 35.7%; Score 614; DB 2; Length 387;
Best Local Similarity 40.8%; Pred. No. 6.6e-35;
Matches 128; Conservative 62; Mismatches 88; Indels 36; Gaps 6;

QY 12 DPWSSONESLSKPYNHLLKPKGNFRNLIVQINRNMVLPKDLQALVSIQVIELLHNS 71
Db 67 DGNWSQVEYKILMGPDYMQHQHKGDIRQLITAFNMVLQVRPESLSITKVVGLHTAS 126
QY 72 LLIDIEDNAPLRGGTSHLIFGVPSSTINTANYMFRAMOLVSQLTKEPLVHNLITIF 131
Db 127 LLVDDVEDNSILRRGIPVAHNIFGTAQTINSANYVFLALQEVQKLNPP-----TAIDIF 181
QY 132 NEELNLHRCQGLDIYWRDFPEIIPQEMYLNMVNMKGTGLFRLLRLMEALSPSSHGH 191
Db 182 VQELLNLHRCQGLDFWRDTL--TCPTEDDYLDWVGNKTYGGLFRLLAQLMQAESIT--G 236
QY 192 HSLVFPFNLGLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQ 251
Db 237 IDCVSIVNMVGLIFQICDDYLNLSNKTYTNQKGLCEDLTEGKFSFPVHISIR-----SN 290
QY 252 EQHNEILRLILRTSDKDIKLIQLLEFDNLSAYTKNFQNLVNMKNENKYLPLDL 311
Db 291 PRNHQISILKORTKDEVEKLYAISYME-STGSFAYTRKVVREF----- 333
QY 312 ASHSDTATNLHDEL 325
Db 334 ---RDKALSLIDEI 344

RESULT 12
Q6F6E2
ID Q6F6E2 PRELIMINARY; PRT; 343 AA.
AC Q6F6E2;

```

```

DT 01-OCT-2004 (TREMELrel. 28, Created)
DT 01-OCT-2004 (TREMELrel. 28, Last sequence update)
DT 01-OCT-2004 (TREMELrel. 28, Last annotation update)
DE Geranylgeranyl diphosphate synthase.
GN Name=PbGGPS;
OS Phoma betae.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Phoma.
OX NCBI_TaxID=137527;
RN [1]
RP SEQUENCE FROM N.A.
RA Toyomasu T., Nakaminami K., Toshima H., Mie T., Watanabe K., Ito H.,
RA Matsui H., Mitsuhashi W., Sassa T., Oikawa H.;
RT "Cloning of a Gene Cluster Responsible for the Biosynthesis of
RT Diterpene Aphidicolin, a Specific Inhibitor of DNA Polymerase alpha.";
RL Biosci. Biotechnol. Biochem. 68:146-152(2004).
DR EMBL; AB079897; BAD29965.1; -
SQ SEQUENCE 343 AA; 38975 MW; 5C873525B973CB50 CRC64;

Query Match 35.6%; Score 612.5; DB 2; Length 343;
Best Local Similarity 43.0%; Pred. No. 7.2e-35;
Matches 123; Conservative 60; Mismatches 86; Indels 17; Gaps 6;

QY 15 WSSQNESLISKPYNHLLKPGKFRNLIVQINRVNMLPKDLAIVSQIVVELLHNSLLI 74
DB 48 WTSTKEKVVTPGYDYIAAKPGKEVRTLLACFDEWLQVPPSPESLEVIGQVVRMLHTASLLI 107

QY 75 DDIEDNAPLRGQTTSHLIFGVPTSTINTANYMYFRAMQLVSLTTPKEPLYHNLITIFNEE 134
DB 108 DDIQNSLRLRGKPAQNIQFALTINSANYVYFLALEKINSL--KNP---NITDIFTEE 162

QY 135 LINLRGGLDIYWRDFLPEIPTQEMYNLMVMNKTGGLFRLTLRLMEALSPSSHGHSL 194
DB 163 LLRLHRGQAMDLYRDTL--TCPTEEYFEWVANNVTGGLFWLMYRMMA--ESSMPIDL 217

QY 195 VPFNLGLIYQIRDDYLNKDFQMSSEKGAEDITEGKLSPPIVHALNFTKTKGQTEQH 254
DB 218 LPVVELLGVIFQVDDYKNCRSRYGKLGKGFEDLTEGKFSFPVHSIR-----SNPED 271

QY 255 NEILRLRLTSDDIKLKLQILEFDNSLAYTKNFQINQVNMVK 300
DB 272 LQLLHVLOQKSSNEHVKLYAIEIME-STGSLEYTKHVVENIVSQIQ 316

Query Match 34.8%; Score 598.5; DB 2; Length 343;
Best Local Similarity 42.3%; Pred. No. 6.9e-34;
Matches 121; Conservative 59; Mismatches 89; Indels 17; Gaps 6;

QY 15 WSSQNESLISKPYNHLLKPGKFRNLIVQINRVNMLPKDLAIVSQIVVELLHNSLLI 74
DB 48 WTSTKEKVVTPGYDYIAAKPGKEVRTLLACFDEWLQVPPSPESLEVIGQVVRMLHTASLLI 107

```

```

QY 75 DDIEDNAPLRGQTTSHLIFGVPTSTINTANYMYFRAMQLVSLTTPKEPLYHNLITIFNEE 134
DB 108 DDIQNSLRLRGKPAQNIQFALTINSANYVYFLALEKINSL--KNP---NITDIFTEE 162
QY 135 LINLRGGLDIYWRDFLPEIPTQEMYNLMVMNKTGGLFRLTLRLMEALSPSSHGHSL 194
DB 163 LLRLHRGQAMDLYRDTL--TCPTEEYFEWVANNVTGGLFWLMYRMMA--ESSMPIDL 217
QY 195 VPFNLGLIYQIRDDYLNKDFQMSSEKGAEDITEGKLSPPIVHALNFTKTKGQTEQH 254
DB 218 LPVVELLGVIFQVDDYKNCRSRYGKLGKGFEDLTEGKFSFPVHSIR-----SNPED 271
QY 255 NEILRLRLTSDDIKLKLQILEFDNSLAYTKNFQINQVNMVK 300
DB 272 LQLLHVLOQKSSNEHVKLYAIEIME-STGSLEYTKHVVENIVSQIQ 316

RESULT 14
Q9P885
ID Q9P885 PRELIMINARY; PRT; 303 AA.
AC Q9P885;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Geranylgeranyl pyrophosphate synthase (EC 2.5.1.29).
GN Name=carg;
OS Mucor circinelloides f. lusitanicus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OX Mucor.
OC NCBI_TaxID=29924;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS277.45;
RA MEDLINE=22581058; PubMed=12695851;
RA Velayos A., Papp T., Aguilar-Elena R., Fuentes-Vicente M.,
RA Esalva A.P., Iturriga E.A., Alvarez M.I.;
RT "Expression of the carg gene, encoding geranylgeranyl pyrophosphate
RT synthase, is up-regulated by blue light in Mucor circinelloides.";
RT Curr. Genet. 43:112-120(2003).
DR EMBL; AJ276129; CAB89115.1; -.
DR GO; GO:0004311; F:farnesyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl_synth.
DR Pfam; PF00348; polyprenyl_synth.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; UNKNOWN_1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR Transferase.
SQ SEQUENCE 303 AA; 34787 MW; D440757CA7D0E215 CRC64;

Query Match 34.4%; Score 591.5; DB 2; Length 303;
Best Local Similarity 41.1%; Pred. No. 1.8e-33;
Matches 113; Conservative 67; Mismatches 78; Indels 17; Gaps 5;

QY 20 ESLISKPYNHLLKPGKFRNLIVQINRVNMLPKDLAIVSQIVVELLHNSLLIDIED 79
DB 14 EDILEPFTLLISQPGKDIRAKLISAFDLWLHVPKDVLVCINKITGLHNASLMDVQD 73
QY 80 NAPLRGQTTSHLIFGVPTSTINTANYMYFRAMQLVSLTTPKEPLYHNLITIFNEELNLH 139
DB 74 DSDLRGVPAHLYGVFPQINTANYVIFLALQEVWMLNIP-----SNMQVCTEELNLH 128
QY 140 RGQGLDIYWRDFLPEIPTQEMYNLMVMNKTGGLFRLTLRLMEALSPSSHGHSLVPFIN 199
DB 129 RGQGLIYWRDSL--TCPTEEYIDVMNNTSGLLRLAVRLMQAASESD---IDYTLVN 183
QY 200 LGLIYQIRDDYLNKDFQMSSEKGAEDITEGKLSPPIVHALNFTKTKGQTEQHNLIR 259
DB 184 IIGHFQVRDDYNNLQSTSYNNKGFCEDLTEGKFSFPVHSIR-----KDPNSRLN 237
QY 260 ILLRLTSDDIKLKLQILEFDNSLAYTKNFQINQ 294

```

Db 238 IISQPTSIKVKYALEVIR-KAGSFYRZFLRQ 271

RESULT 15

GGPP_HUMAN STANDARD; PRT; 300 AA.

AC O95749;

AD 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPGASE)

DE (Geranylgeranyl diphosphate synthase) [Includes:

DE Dimethylallyltranstransferase (EC 2.5.1.1); Geranyltranstransferase

DE (EC 2.5.1.10); Farnesyltranstransferase (EC 2.5.1.29)].

GN Name=GGPS1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal heart;

RK MEDLINE=98412715; PubMed=9741684;

RA Ericsson J., Greene J.M., Carter K.C., Shell B.K., Duan D.R.,

RA Florence C., Edwards P.A.;

RT "Human geranylgeranyl diphosphate synthase: isolation of the cDNA,

RT chromosomal mapping and tissue expression.";

RL J. Lipid Res. 39:1731-1739(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RK MEDLINE=99150380; PubMed=10026212;

RA Kuzuguchi T., Morita Y., Segami I., Segami H., Ogura K.;

RT "Human geranylgeranyl diphosphate synthase. cDNA cloning and

RT expression.";

RL J. Biol. Chem. 274:5888-5894(1999).

RN [3]

RP SEQUENCE FROM N.A.

RC Misawa N., Okazaki H., Noguchi Y., Tatsuno I., Saito Y., Yasuda T.,

RA Hirai A.;

RT "Study on isolation of a geranylgeranyl pyrophosphate (GGPP) synthase

RT cDNA and its expression - development of a new assay system of gene

RT functions.";

RL Proc. Jpn. Conf. Biochem. Lipids 41:293-296(1999).

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Placental;

RK MEDLINE=20402571; PubMed=10931946;

RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,

RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,

RA Gu B.-W., Dai N., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,

RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,

RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.;

RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal

RT axis and full-length cDNA cloning.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver, and Spleen;

RK MEDLINE=99203156; PubMed=10101267;

RA Kainou T., Kawamura K., Tanaka K., Matsuda H., Kawamukai M.;

RT "Identification of the GGPS1 genes encoding geranylgeranyl diphosphate

RT synthases from mouse and human.";

RL Biochim. Biophys. Acta 1437:333-340(1999).

RN [6]

RP SEQUENCE FROM N.A.

RC Zhang M., Yu L., Hu P., Bi A., Zhang Q., Xu M., Zhao S.;

RT "Molecular cloning and expression analysis of a novel human cDNA

RT encoding a protein homologous to Neurospora crassa geranylgeranyl

RT pyrophosphate synthetase.";

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RK MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldi A.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.J., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.R.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of

CC IPP onto DMAPP to form geranylgeranyl pyrophosphate, an important

CC precursor of carotenoids and geranylated proteins.

CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl

CC diphosphate = diphosphate + geranyl diphosphate.

CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate

CC = diphosphate + trans,trans-farnesyl diphosphate.

CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl

CC diphosphate = diphosphate + geranylgeranyl diphosphate.

CC -!- PATHWAY: Isoprenoid biosynthesis.

CC -!- SUBUNIT: Homooctamer.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- TISSUE SPECIFICITY: Abundantly expressed in testis. Found in other

CC tissues to a lower extent.

CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; AB017971; BAA75909.1; -

CC EMBL; AB019036; BAA77251.1; -

CC EMBL; AF125394; AAD43050.1; -

CC EMBL; AB016043; BAA76511.1; -

CC EMBL; AF057698; AAG45581.1; -

CC EMBL; BC005252; AAH05252.1; -

CC Genew; HGNC:4249; GGPS1.

CC MIM; 605982; -

CC GO; GO:0004337; F:geranyltranstransferase activity; TAS.

CC GO; GO:0006629; P:lipid metabolism; TAS.

CC InterPro; IPR000092; Polyprenyl synth.

CC InterPro; IPR008949; Terpenoid synth.

CC Pfam; PF00348; Polyprenyl synth. 1.

CC PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.

CC PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.

CC Isoprene biosynthesis; Transferase.

CC SEQUENCE 300 AA; 34871 MW; F5D1959274BEE27A CRC64;

Query Match 33.9%; Score 583; DB 1; Length 300;

Best Local Similarity 40.5%; Pred. No. 7.1e-33;

Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;

QY 20 ESLISKPYNHILKPGKFNFLNLIVQINRMVNLPKDLAIQSVQIVELLNSSLIDIED 79

Db 9 QRILLEPYKYLQLPQKQVTKLSQAFNHLKVPEDKQLIIIEVTMLNHLASLLIDIED 68

QY 80 NAPLRGQTTSHLIFGVPSTINTANTYVYFAMQIVSQLTTPKEPLYHNLITFNEELINLH 139

| | | | |
|----|-----|---|-----|
| Db | 69 | NSKLRGPFVAHSIYGIPSVINSANYVYFLGEKV--LTIIDHP--DAVKLFTRQLLEIH | 123 |
| Qy | 140 | RGQGLDIYWRDPLPEIIPQEMYLNMVNMNKTGGLFELTLRLMEALSPSSHGHGSLVPPFIN | 199 |
| Db | 124 | QOQGLDIYWRD--NYTCPEEYKAMVLOKTGGLFGLAVGLMQLP---SDYKEDLKPLIN | 178 |
| Qy | 200 | LLGIYQIRDDYINLKDQMSSEKGFADITREKLSFFPIVHALNFTKTKGQTEQHNEIIR | 259 |
| Db | 179 | TLGLFFQIRDDYANLHSEKSEKSFCELTGKFSFPTIHAI-----WSRPESTQVQN | 232 |
| Qy | 260 | ILLRTSDKDIKLIQILEFTNSLAYTKNFINOIVNMKNENKYLPLDLASHSDTAT | 319 |
| Db | 233 | ILRQFTENIDIKKCVHYLE-DVGSFEXTRNTLKEI-----EAK-----AYKQIDAR | 278 |
| Qy | 320 | NLHDELLYIIDHLSL | 335 |
| Db | 279 | GGNPELVALVKELSKM | 294 |

Search completed: October 17, 2004, 08:52:57
Job time : 125.839 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:05:06 ; Search time 21.8543 Seconds
(without alignments)
1474.884 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINNDPWSQNE.....DTATNLHDELLYIIDHLSL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 1720 | 100.0 | 335 | 2 | farnesyltranstrans |
| 2 | 623 | 36.2 | 428 | 2 | farnesyltranstrans |
| 3 | 337 | 19.6 | 347 | 2 | octaprenyl-diphosp |
| 4 | 310.5 | 18.1 | 327 | 2 | bifunctional short |
| 5 | 274.5 | 16.0 | 364 | 2 | probable octaprenyl |
| 6 | 244 | 14.2 | 325 | 2 | bifunctional short |
| 7 | 243.5 | 14.2 | 323 | 2 | octaprenyl-diphosp |
| 8 | 239 | 13.9 | 325 | 2 | bifunctional short |
| 9 | 238.5 | 13.9 | 324 | 2 | octaprenyl-diphosp |
| 10 | 233 | 13.5 | 322 | 2 | octaprenyl-diphosp |
| 11 | 228.5 | 13.3 | 330 | 2 | farnesyltranstrans |
| 12 | 226 | 13.1 | 323 | 2 | octaprenyl-diphosp |
| 13 | 223 | 13.0 | 323 | 2 | octaprenyl-diphosp |
| 14 | 223 | 13.0 | 323 | 2 | octaprenyl-diphosp |
| 15 | 223 | 13.0 | 323 | 2 | octaprenyl-diphosp |
| 16 | 214.5 | 12.5 | 332 | 2 | octaprenyl-diphosp |
| 17 | 211.5 | 12.3 | 332 | 2 | octaprenyl-diphosp |
| 18 | 210.5 | 12.2 | 330 | 2 | hypothetical prote |
| 19 | 210.5 | 12.2 | 332 | 2 | hypothetical prote |
| 20 | 209.5 | 12.2 | 346 | 2 | farnesyltranstrans |
| 21 | 205 | 11.9 | 323 | 2 | bifunctional short |
| 22 | 205 | 11.9 | 323 | 2 | prephytoene pyroph |
| 23 | 204 | 11.9 | 329 | 2 | probable prenyl tr |
| 24 | 203 | 11.8 | 297 | 2 | hypothetical prote |
| 25 | 202 | 11.7 | 348 | 2 | polyprenyl synthet |
| 26 | 201 | 11.7 | 299 | 2 | octaprenyl-diphosp |
| 27 | 198 | 11.5 | 342 | 2 | hypothetical prote |
| 28 | 197.5 | 11.5 | 338 | 2 | probable geranylge |
| 29 | 195.5 | 11.4 | 304 | 2 | polyprenyl synthet |
| | | | | | octoprenyl-diphosp |

| | | | | | | |
|----|-------|------|-----|---|--------|---------------------|
| 30 | 193.5 | 11.2 | 371 | 2 | D97790 | octaprenyl-diphosp |
| 31 | 193 | 11.2 | 337 | 2 | B72560 | probable geranyltr |
| 32 | 192.5 | 11.2 | 327 | 2 | D71651 | octaprenyl-diphosp |
| 33 | 191.5 | 11.1 | 326 | 2 | F69535 | geranylgeranyl dip |
| 34 | 191 | 11.1 | 362 | 2 | AF3443 | geranyltranstransf |
| 35 | 188.5 | 11.0 | 281 | 2 | A93405 | hypothetical prote |
| 36 | 187.5 | 10.9 | 319 | 2 | F86792 | hypothetical prote |
| 37 | 186.5 | 10.8 | 338 | 2 | AD2654 | octaprenyl-diphosp |
| 38 | 186.5 | 10.8 | 358 | 2 | H97435 | decaprenyl diphosp |
| 39 | 185 | 10.8 | 348 | 2 | B69630 | heptaprenyl diphosp |
| 40 | 184 | 10.7 | 272 | 2 | G72410 | geranyltranstransf |
| 41 | 182.5 | 10.6 | 323 | 2 | AH1818 | heptaprenyl diphosp |
| 42 | 176 | 10.2 | 326 | 2 | AH1780 | probable heptapren |
| 43 | 176 | 10.2 | 335 | 2 | E70549 | trans-pentaprenyl |
| 44 | 175.5 | 10.2 | 473 | 1 | XUBYTP | heptaprenyl diphos |
| 45 | 172 | 10.0 | 326 | 2 | A11404 | heptaprenyl diphos |

ALIGNMENTS

RESULT 1

S60921

farnesyltranstransferase (EC 2.5.1.29) - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: geranylgeranyl-diphosphate synthase; protein lPE1c; protein YPL069c
C:Species: *Saccharomyces cerevisiae*

C>Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004

C/Accession: S60921; S61936

R:Winnett, E.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;
submitted to the EMBL Data Library, October 1995

A:Reference number: S60921

A:Accession: S60921

A:Molecule type: DNA

A:Residues: 1-335 <WIN>

A:Cross-references: UNIPROT:Q12051; EMBL:U39205; NID:gi079672; PIDN:AAB68296.1; PID:gl0797;
R:Jiang, Y.; Proteau, P.; Poulter, D.; Ferro-Novick, S.

submitted to the EMBL Data Library, July 1995

A:Description: Bts1 encodes a geranylgeranyl diphosphate synthase in *Saccharomyces cerevisiae*

A:Reference number: S61936

A:Accession: S61936

A:Molecule type: DNA

A:Residues: 1-335 <JIA>

A:Cross-references: EMBL:U31632; NID:gl098640; PIDN:AAA83262.1; PID:gl098641

C:Genetics:

A:Gene: SGD:BTS1

A:Cross-references: SGD:S0005990; MIPS:YPL069c

A:Map position: 16L

C:Superfamily: prenyl transferase A

C:Keywords: transferase

Query Match 100.0%; Score 1720; DB 2; Length 335;
Best local similarity 100.0%; Pred. No. 4.3e-114;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | MEAKIDELINNDPWSQNESLSKPNHLLKPGKFRNLIVQINRVNVLPRDQLAIV | 60 |
| DB | 1 | MEAKIDELINNDPWSQNESLSKPNHLLKPGKFRNLIVQINRVNVLPRDQLAIV | 60 |
| QY | 61 | SQIVELHNSLLIDIEDNAPLRGQTTSLLIFGVPSITNTANTMYFRAMQLVSLTTK | 120 |
| DB | 61 | SQIVELHNSLLIDIEDNAPLRGQTTSLLIFGVPSITNTANTMYFRAMQLVSLTTK | 120 |
| QY | 121 | EPIYHNLITIFNEELINLHARGQGLDIYWRDFFPEIIPTEQMYLNMMVNTKGLFRLLRL | 180 |
| DB | 121 | EPIYHNLITIFNEELINLHARGQGLDIYWRDFFPEIIPTEQMYLNMMVNTKGLFRLLRL | 180 |
| QY | 181 | MEALSPSSHGHSLVPFINLGIYQIRDDYLNKDPQMSSEKGFADITEGKLSFPIVH | 240 |
| DB | 181 | MEALSPSSHGHSLVPFINLGIYQIRDDYLNKDPQMSSEKGFADITEGKLSFPIVH | 240 |
| QY | 241 | ALNFTKTKGQTEQHNEILRLITSDKDKLKLIQILEFDNTSLAYTKNFNQLVNMIR | 300 |
| DB | 241 | ALNFTKTKGQTEQHNEILRLITSDKDKLKLIQILEFDNTSLAYTKNFNQLVNMIR | 300 |

QY 301 NDNENKYLPLASHSDTATNLHDELLYIIDLSEL 335
 Db 301 NDNENKYLPLASHSDTATNLHDELLYIIDLSEL 335

RESULT 2
 S15662
 A:Title: farnesyltransferase (EC 2.5.1.29) [validated] - Neurospora crassa
 N:Alternate names: geranylgeranyl pyrophosphate synthetase
 C:Species: Neurospora crassa
 C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
 C:Accession: S15662; T46598; T46592
 R:Carattoli, A.; Romano, N.; Ballarino, P.; Morelli, G.; Macino, G.
 J. Biol. Chem. 266, 5854-5859, 1991
 A:Title: The Neurospora crassa carotenoid biosynthetic gene (albino 3) reveals highly co
 A:Reference number: S15662; MUID:91170267; PMID:1826006
 A:Accession: S15662
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-428 <CAR>
 A:Cross-references: UNIPROT:P24322; GB:U20940; EMBL:X53979; NID:G903318; PIDN:ACU3867.1
 R:Vittorio, P.; Carattoli, A.; Londei, P.; Macino, G.
 J. Biol. Chem. 269, 26650-26654, 1994
 A:Title: Internal translational initiation in the mRNA from the Neurospora crassa albino
 A:Reference number: A50065; MUID:95014519; PMID:7929398
 A:Accession: S15662
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-129 <VIT1>
 A:Cross-references: EMBL:S74011; PIDN:CAB33185.1
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 'MEHVT', 1-129 <VIT1>
 A:Cross-references: EMBL:S74011; PIDN:CAB33185.1
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 'MEHVT', 1-77, 'PACQLSPRHTOWRPPSPRLPTQTLRPK', 'TSSAPRGAPGPKRRRR' <VIT2>
 A:Cross-references: EMBL:S74685; PIDN:CAB33252.1
 A:Experimental source: mutant rosv
 A:Note: translation starts at an alternate initiator and produces a partially active enz
 C:Genetics:
 A:Gene: al-3
 C:Keywords: transferase

Query Match 36.2%; Score 623; DB 2; Length 428;
 Best Local Similarity 39.1%; Pred. No. 1.5e-36;
 Matches 125; Conservative 74; Mismatches 99; Indels 22; Gaps 7;

QY 15 WSSQNESLSKPNHLLKPKGNFRNLIVQINRVNMLPKDQIAIVSQVELHNSLLI 74
 Db 121 WSEKEKVLTPYDYLNGHPKDIRSQMKAFDAMLDPSESLVITKVISMLHTASLLV 180

QY 75 DIEDNAPLRGQTTSHLIFGVPTSTNTANYMYFRAMOLVSQLTTEKPLVHNLITIFNER 134
 Db 181 DVEDNSVLRGFPVAHSIFGIPOTINTSNVYVYFALQELQK--KNP---KAVSIFSEE 235

QY 135 LNLHRRGQGLDIYWRDPLPEIITQEMYLNNVNNKGGFLFRLLRLMEALSPSSHGHSL 194
 Db 236 LNLHRRGQGLDIYWRDPL--TCETEDDYLEWVSKTKGGLFRIGIKLMAQESRSP---VDC 290

QY 195 VPFNLGLIYQTRDDYLNKDFQMSSEKGFADITEGKLSFPIVIALNFTKGTQEQH 254
 Db 291 VPLVNIIGLTFQADDYHNLWREYANKGMCDELTEGKSFVFIHSI-----SNPSN 344

QY 255 NEILRIILLTSDKDKLQILFPTNSLATYKFNINLVNMIKNDNENKYLPLASH 314
 Db 345 MQLNLILKQTKGDEEVARYAYME-STGSFETIRKVKVLV-----DRARQMTEDIDGG 398

QY 315 SDTATNLHDELLYIIDLSEL 334
 Db 399 RGSKGTHKILDRIMLHQEE 418

RESULT 3
 A70139

octaprenyl-diphosphate synthase (ispB) homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Sep-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
 C:Accession: A70139
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: A70139
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-347 <KLE>
 A:Cross-references: UNIPROT:O51294; GB:AE001138; GB:AE000783; NID:G2688210; PIDN:AAC6669
 A:Experimental source: strain B31
 C:Superfamily: prenyl transferase A

Query Match 19.6%; Score 337; DB 2; Length 347;
 Best Local Similarity 29.9%; Pred. No. 1.7e-16;
 Matches 92; Conservative 57; Mismatches 125; Indels 34; Gaps 7;

QY 23 ISKPNHLLKPKGNFRNLIVQINRVNMLPKDQIAIV---SQVELHNSLLIDDED 79
 Db 46 IKAPAEIIVNRGGKRIEPMIMILLAYALGKKNKLIYKUSLLELPHSGSLIDDED 105

QY 80 NAPLRGQTTSHLIFGVPTSTNTANYMYFRAMOLVSQLTTEKPLVHNLITIFNEEL 135
 Db 106 NSLKRGSALHLLIYGDINSAGNLIYFLPAKLEKSNLKNQKLLIYENFT---TL 161

QY 136 LNLHRRGQGLDIYWRDPLPEIITQEMYLNNVNNKGGFLFRLLRLMEALSPSSHGHSLV 195
 Db 162 SNLHGGQIDIKFHN--ESYIPSIKEYISLVELKTASLFGMASFLAAILTNNEDEKAKKIY 219

QY 196 PFNILLGLIYQTRDDYLNKDFQMSSEKGFADITEGKLSFPIVIALNFTKGTQEQH 242
 Db 220 STFLKLGVPYFQIIDDIKIKN--KINGKEFGDLDLEKKSLFIIFLQEKKEPKIISKF 277

QY 243 NFKTKGQTEQHNELIRLLTSLTSDKDIK-----LKLIFILEFTNSLAYTKNFINQLV 296
 Db 278 NQIKNTKTKARKEIFKLTEMINSKSIKNSITVALKYLNEFKNELNLYPLTNRYKLLI 337

QY 297 NMKNDNE 304
 Db 338 DTIEQIKE 345

RESULT 4
 D64407
 bifunctional short chain isoprenyl diphosphate synthase homolog - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
 C:Accession: D64407
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
 Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.,
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: D64407
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-327 <BUL>
 A:Cross-references: UNIPROT:Q58270; GB:U67530; GB:I77117; NID:G2826340; PIDN:AAB98865.1
 C:Genetics:
 A:Map position: REV783844-782861
 A:Start codon: TTG
 C:Superfamily: prenyl transferase A

Query Match 18.1%; Score 310.5; DB 2; Length 327;
 Best Local Similarity 27.6%; Pred. No. 1.2e-14;

RESULT 6

T45152
C;Species: Methanobacterium thermoautotrophicum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C;Accession: T45152
R;Chen, A.; Poulter, C.D.
A;Title: Biochem. Biophys. 314, 399-404, 1994
A;Title: Isolation and characterization of idsA: the gene for the short chain isoprenyl diphosphate synthase [imported] - Methanobacterium thermoautotrophicum
A;Reference number: Z22932; MUID:95070145; PMID:7979381
A;Accession: T45152
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-325 <CHE>
A;Cross-references: EMBL:S75695; PIDN:AAB32421.1
C;Genetics:
A;Gene: idsA
C;Superfamily: prenyl transferase A

| | Query Match | 14.2%; | Score 244; | DB 2; | Length 325; |
|---------|---|------------------|--------------------|------------|-------------|
| | Best Local Similarity | 26.3%; | Pred. No. 5.6e-10; | | |
| Matches | 82; | Conservative 61; | Mismatches 145; | Indels 24; | Gaps 10; |
| QY | 25 KPNHILKPKGNFLNL-IIVQINVVMNLPKDQLAIVSQIIVELLHNSSLIDIEDINAPL 83 | : | : | : | : |
| Db | 33 KASEHLITAGCKIRPSLLALSCEAVGGNPED-AAGVAAAIELIHFTSLIHDDIMDDDEM 91 | : | : | : | : |
| QY | 84 RRGOTTSHLI FGVPSTINTANVMYFRAMQLVSQLTTKPELYHNLIITIFNEELINLHRGQG 143 | : | : | : | : |
| Db | 92 RRGPSPVHVIMGEPMAILLAGDLVLFSKAFAEVIIRNGDSERV-KDALAVVDVSCVKICGQA 150 | : | : | : | : |
| QY | 144 LDIIYWRDFLPBIIITQEMYLNMMVNKTGLFRLTILRMALSPSSHIG-HSLVPFFINLG 202 | : | : | : | : |
| Db | 151 LDM--GFEEERLDVTEDEYMENIYKTAALIAATKAGAIMGCASREVEALEDYGKF IG 207 | : | : | : | : |
| QY | 203 IIVQIRDYLMLKDPQMSESGFAEDIETGKUSFPFI VHAFNIYTKTGQTQHNEILRI LL 262 | : | : | : | : |
| Db | 208 LAFOIHDDYLDVWSDEESIGPVGSDIAEGKMTLMVVKAL-----BEASEDRERL- ISI 261 | : | : | : | : |
| QY | 263 LRISDKDIKLKIIGILEPTDTNSLA YTKNFINOLVNMKNNDENKYL PDLASHSDYTATNLH 322 | : | : | : | : |
| Db | 262 LGSGDEGSVAEAIEIFE-RYGATQYAHEVALDYVRMAKERLE-----ILEDSDA----R 310 | : | : | : | : |
| QY | 323 DELLYYIIDHLSE 334 | : | : | : | : |
| Db | 311 DALMRIADFVLE 322 | : | : | : | : |

```

RESULT 7
AB0427
octaprenyl-diphosphate synthase (EC 2.5.1.-) [imported] - Yersinia pestis (strain C092)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AB0427
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0427
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <KUR>
A;Cross-references: UNIPROT:Q8ZE45; GB:AL590842; PID:CAC92742.1; PID:gl5981437; GSPDB:GN
C;Genetics:
A;Gene: ispB
C;Superfamily: prenyl transferase A
C;Keywords: transferase

Query Match          14.2%; Score 243.5; DB 2; Length 323;
Best Local Similarity 24.8%; Pred. No. 6e-10;

```


| No. | Score | Match | Length | DB | ID | Description |
|-----|-------|-------|--------|----|---------------------|-------------|
| 1 | 1720 | 100.0 | 335 | 14 | US-10-369-493-22494 | Sequence 2 |
| 2 | 1720 | 100.0 | 335 | 15 | US-10-450-941-6 | Sequence 6 |
| 3 | 1720 | 100.0 | 335 | 15 | US-10-041-018-22 | Sequence 1 |
| 4 | 629.5 | 36.6 | 348 | 14 | US-10-369-493-12405 | Sequence 2 |
| 5 | 623 | 36.2 | 428 | 14 | US-10-369-493-4221 | Sequence 4 |
| 6 | 591.5 | 34.4 | 303 | 15 | US-10-041-018-34 | Sequence 3 |
| 7 | 583 | 33.9 | 300 | 15 | US-10-041-018-46 | Sequence 4 |
| 8 | 577 | 33.5 | 304 | 9 | US-09-925-302-614 | Sequence 61 |
| 9 | 577 | 33.5 | 304 | 10 | US-09-925-302-614 | Sequence 6 |
| 10 | 575.5 | 33.5 | 452 | 14 | US-10-369-493-12962 | Sequence 1 |
| 11 | 545.5 | 31.7 | 338 | 15 | US-10-041-018-24 | Sequence 2 |
| 12 | 545.5 | 31.7 | 338 | 15 | US-10-041-018-77 | Sequence 7 |
| 13 | 538 | 31.3 | 376 | 15 | US-10-001-192A-15 | Sequence 1 |
| 14 | 513 | 29.8 | 335 | 15 | US-10-041-018-23 | Sequence 2 |

QY 121 EPLYHNLTIFNEELINLHRGQGLDIYWRDFLPEIIPTEMYLNMVMNKTGGLFRLTLRL 180

SUMMARIES

Db 121 EPLVHNLITIFNEELINLHROGGLDIYWRDPLPEIIPQEMVYNMVKTKGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
QY 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMVK 300
Db 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMVK 300
QY 301 NDENKYLPLDASHSDTATNLHDELYIIDHLSL 335
Db 301 NDENKYLPLDASHSDTATNLHDELYIIDHLSL 335

RESULT 2

US-10-450-941-6
; Sequence 6, Application US/10450941
; Publication No. US20040063182A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: A METHOD OF PRODUCING PRENYLALCOHOL
; FILE REFERENCE: PH-1444PCT
; CURRENT APPLICATION NUMBER: US/10/450,941
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: JP2000-401701
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2000-403067
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2001-282978
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 6
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-450-941-6

Query Match 100.0%; Score 1720; DB 15; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.1e-140;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
Db 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
Db 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
QY 121 EPLVHNLITIFNEELINLHROGGLDIYWRDPLPEIIPQEMVYNMVKTKGGLFRLTLRL 180
Db 121 EPLVHNLITIFNEELINLHROGGLDIYWRDPLPEIIPQEMVYNMVKTKGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
QY 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMVK 300
Db 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMVK 300
QY 301 NDENKYLPLDASHSDTATNLHDELYIIDHLSL 335
Db 301 NDENKYLPLDASHSDTATNLHDELYIIDHLSL 335

RESULT 3

US-10-041-018-22
; Sequence 22, Application US/10041018
; Publication No. US20040072323A1

; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-22

Query Match 100.0%; Score 1720; DB 15; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.1e-140;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
Db 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
Db 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
QY 121 EPLVHNLITIFNEELINLHROGGLDIYWRDPLPEIIPQEMVYNMVKTKGGLFRLTLRL 180
Db 121 EPLVHNLITIFNEELINLHROGGLDIYWRDPLPEIIPQEMVYNMVKTKGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
QY 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMVK 300
Db 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMVK 300
QY 301 NDENKYLPLDASHSDTATNLHDELYIIDHLSL 335
Db 301 NDENKYLPLDASHSDTATNLHDELYIIDHLSL 335

RESULT 4

US-10-369-493-12405
; Sequence 12405, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12405
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-369-493-12405

Query Match 36.6%; Score 629.5; DB 14; Length 348;
Best Local Similarity 43.2%; Pred. No. 5.2e-46;
Matches 127; Conservative 63; Mismatches 87; Indels 17; Gaps 6;

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 473/4
SEQ ID NO 12982
LENGTH: 452
TYPE: PRT
ORGANISM: Aspergillus nidulans
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(452)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12982

Query Match 33.5%; Score 575.5; DB 14; Length 452;
Best Local Similarity 40.3%; Pred. No. 3.5e-41;
Matches 133; Conservative 47; Mismatches 99; Indels 51; Gaps 6;

QY 10 NNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNMNLPKQDLAIYSQIVELLHNS 69
DB 105 DNOHQHKEEIIISAPVHYISTLPGKDIRGKLISAFNEWFQIPDEQLEIKRAVGLHV 164
QY 70 SSL-----IDDEDNAPLRGQTTSHLIFGVPSTINTANYM 106
DB 165 ASLLKGSFTLTHQITQFRQDNLCTRRIDDIEDYKLRGFPVAHSIFGIPQTINSANYA 224
QY 107 YFRAMOLVSQITTEPELYVHNLITFNEELINHRGQGLDIYWRDPLPFIITQENYLMV 166
DB 225 YFOAQNEVLKLRCE---KALRIFTTELLHRGQGMIDYWRDSL--TCPSEBYLDMV 278
QY 167 MNKGGGLFRLRLMEALSPSSHGHSI-----VPEINLIGIYO 206
DB 279 ANKGGGLFRLAIKLIQLESVDVEFVRPLFELPLFTRAEQYKLRRDCVPLVDLGIIFQ 338
QY 207 IRDYLNLKDFQMSSEKGFADITEGKLSFPFIVHALNFTTKG-QTEOHNEILRLILRT 265
DB 339 IRDDYQNLQSDQYAKNKGFGEDITEGKFSYPIVHSIRSSRTGVSQSSQLQLILRQKT 398
QY 266 SDKIKLKLQILFEDNLSLAYTKNFINQL 295
DB 399 BDEAVKXYTIQILR-KTGSFEFTKQLRDL 427

RESULT 11
US-10-041-018-24
Sequence 24, Application US/10041018
Publication No. US20040072323A1
GENERAL INFORMATION:
APPLICANT: Matsuda, Seiichi P.T.
APPLICANT: Hart, Elizabeth A.
TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
FILE REFERENCE: P02080US/10025547
CURRENT APPLICATION NUMBER: US/10/041,018
CURRENT FILING DATE: 2002-01-07
PRIOR APPLICATION NUMBER: US 60/259880
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 413
SOFTWARE: Patent in version 3.1
SEQ ID NO 24
LENGTH: 338
TYPE: PRT
ORGANISM: Drosohila melanogaster
US-10-041-018-24

Query Match 31.7%; Score 545.5; DB 15; Length 338;
Best Local Similarity 38.7%; Pred. No. 9.3e-39;
Matches 123; Conservative 55; Mismatches 113; Indels 27; Gaps 7;
QY 12 DPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNMNLPKQDLAIYSQIVELLHNS 71
DB 13 DKSTQKEQDEILLQPFYIQQIPKQFRSELALAFNHLLIPGKLAQIGDIVOMLHNS 72
QY 72 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMOLVSQITTEPELYVHNLITIF 131
DB 73 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMOLVSQITTEPELYVHNLITIF 131
QY 132 NEELINHRGQGLDIYWRDPLPFIITQENYLMNMTKGGELRLTLRLMEALSPSSHG 191
DB 128 TEQLLEHRRGQGMELIYWRDSF--TCPSESDYKLMTRVKTGGLFMAIRLMOLFSSNKEDY 185
QY 192 HSLVFFINLGIYQIRDDYLNLDKDFQMSSEKGFADITEGKLSFPFIVHALNFTTKGQT 251
DB 186 SKLTA---ILGLYFQIRDDYCNLSLKEYTENKSPAEDITEGKFGFPVHAVRTOK----- 237
QY 252 EQHNEILRLILRTSDKDKIKLQILFEDNLSLAYTKNFINQLVNMINKND-----NENKY 307
DB 238 -QDKQVLHILRQTHDIEVKYCTITLE-KLGSFYTRKLVESLDAEARESEVARLGSNPY 295
QY 308 LPDLASH-----SDTAT 319
DB 296 MDRLLKLLSWKTSDSAS 313

RESULT 12
US-10-041-018-77
Sequence 77, Application US/10041018
Publication No. US20040072323A1
GENERAL INFORMATION:
APPLICANT: Matsuda, Seiichi P.T.
APPLICANT: Hart, Elizabeth A.
TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
FILE REFERENCE: P02080US/10025547
CURRENT APPLICATION NUMBER: US/10/041,018
CURRENT FILING DATE: 2002-01-07
PRIOR APPLICATION NUMBER: US 60/259880
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 413
SOFTWARE: Patent in version 3.1
SEQ ID NO 77
LENGTH: 338
TYPE: PRT
ORGANISM: Drosohila melanogaster
US-10-041-018-77

Query Match 31.7%; Score 545.5; DB 15; Length 338;
Best Local Similarity 38.7%; Pred. No. 9.3e-39;
Matches 123; Conservative 55; Mismatches 113; Indels 27; Gaps 7;
QY 12 DPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNMNLPKQDLAIYSQIVELLHNS 71
DB 13 DKSTQKEQDEILLQPFYIQQIPKQFRSELALAFNHLLIPGKLAQIGDIVOMLHNS 72
QY 72 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMOLVSQITTEPELYVHNLITIF 131
DB 73 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMOLVSQITTEPELYVHNLITIF 131
QY 132 NEELINHRGQGLDIYWRDPLPFIITQENYLMNMTKGGELRLTLRLMEALSPSSHG 191
DB 128 TEQLLEHRRGQGMELIYWRDSF--TCPSESDYKLMTRVKTGGLFMAIRLMOLFSSNKEDY 185
QY 192 HSLVFFINLGIYQIRDDYLNLDKDFQMSSEKGFADITEGKLSFPFIVHALNFTTKGQT 251
DB 186 SKLTA---ILGLYFQIRDDYCNLSLKEYTENKSPAEDITEGKFGFPVHAVRTOK----- 237
QY 252 EQHNEILRLILRTSDKDKIKLQILFEDNLSLAYTKNFINQLVNMINKND-----NENKY 307
DB 238 -QDKQVLHILRQTHDIEVKYCTITLE-KLGSFYTRKLVESLDAEARESEVARLGSNPY 295

```
QY 308 LPDLASH-----SDTAT 319
Db 296 MDRLNKLKLSWKTSDSAS 313

RESULT 13
US-10-001-192A-15
; Sequence 15, Application US/10001192A
; Publication No. US20040091958A1
; GENERAL INFORMATION:
; APPLICANT: Ooijen, Albert
; APPLICANT: Verdoes, Jan
; APPLICANT: Wery, Jan
; TITLE OF INVENTION: IMPROVED METHODS FOR TRANSFORMING
; TITLE OF INVENTION: PHAFFIA STRAINS, TRANSFORMED PHAFFIA STRAINS SO OBTAINED AND
; TITLE OF INVENTION: RECOMBINANT DNA IN SAID METHODS
; FILE REFERENCE: 24615-20104.01
; CURRENT APPLICATION NUMBER: US/10/001,192A
; CURRENT FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: EP 95203620.0
; PRIOR FILING DATE: 1995-12-22
; PRIOR APPLICATION NUMBER: EP96200943.7
; PRIOR FILING DATE: 1996-04-11
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Phaffia rhodozyma
US-10-001-192A-15

Query Match 31.3%; Score 538; DB 15; Length 376;
Best Local Similarity 31.1%; Pred. No. 4.8e-38;
Matches 117; Conservative 79; Mismatches 110; Indels 70; Gaps 7;

QY 15 WSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLAIYSQIVVELHNSSLTI 74
Db 14 FTQDDIVLLEPVHYLGKPGKEIRSQLEAFNYWLDVKKEDLEVIQNVVGLHTASLGM 73

QY 75 DIEDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLV-----SDTAT 114
Db 74 DDVEDSVLRGSPVAHLIYGPQTINTANYVFLAYQEIFKLRPTPIPMVIPPSSASL 133

QY 115 -----SQTTEKPEYHNLITFNEELNLRHGQGLDIYWRD 150
Db 134 QSVSVSASSSSASSSNGGTSTPNQIPFSKDTY--LDKVTDEMLSLRGQGLLEFWRD 191

QY 151 FLPEIIPQEMLYNMVNMKTTGGLFRLTLRLMEALSPSHHGLSVLPFINLGLIYQIRD 210
Db 192 SL--TCPSEHYKVLKGTGGLFRIAVRLMA--KSECDIDFVQLVNLISYFQIRDD 246

QY 211 YNLKDFQMSSEKGFADITEGKLSPPVHVALNFTKGTQTEQHNEILRIILLRTSKDI 270
Db 247 YNMLQSSEYAHNKNFAEDITEGKFSPTIHSIHANPSS-----RLVINTLQKKSPEI 300

QY 271 KLKLIQLEFDTNSLAYTKNFQINQVNMKND-----NENKYLPLDLASHSDTAT 319
Db 301 LHCVMYMTTETHSFETQVIMTLNLSGALERELGRQGEFAEANSRMDLGDVDSEGTGK 360

QY 320 NLHDELLYIIDLHSEL 335
Db 361 NV--KLEAILKLLADI 374

RESULT 14
US-10-041-018-23
; Sequence 23, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
```

```
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-041-018-23

Query Match 29.8%; Score 513; DB 15; Length 335;
Best Local Similarity 37.7%; Pred. No. 5.9e-36;
Matches 120; Conservative 56; Mismatches 112; Indels 30; Gaps 8;

QY 12 DPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLAIYSQIVVELHNS 71
Db 13 DKSTQKEQDEILLQPTTYIQIPGKQFRSELALAFNHNLLIPGEKLAQIGNIVQMLHNS 72

QY 72 LIIDDIEDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLVSOLTTEKPLYHNLITIF 131
Db 73 LL---IEDNSILRRGVPAHSIYGVASTINAAAYALFLALEKVQQDDHEF-----VRKY 124

QY 132 NEELINLRHGQGLDIYWRDLPPEIIPQEMLYNMVNMKTTGGLFRLTLRLMEALSPSHHG 191
Db 125 TEQLLELRHGQGLDIYWRDSF--TCPSESDYKLVMTVRKTGGLFRLTLRLMEALSPSHHG 182

QY 192 HSLVFPINLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSPPVHVALNFTKGTQ 251
Db 183 SKLTA---ILGLYFQIRDDYCNLSKEVSNVKSFAEDITEGKFGFPGFVIHAVRTQK----- 234

QY 252 EOHNETLRLILRTSDKDIKLIQLEFDTNSLAYTKNFQINQVNMKND-----NENKY 307
Db 235 -QKQVLRKSSORTHDIEVKYCIITLLE-KLGSFQYTRKVLSELDAAEASSEVSTLGSNPY 292

QY 308 LPDLASH-----SDTAT 319
Db 293 MDRLNKLKLSWKTSDSAS 310

RESULT 15
US-10-041-018-76
; Sequence 76, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-041-018-76

Query Match 29.8%; Score 513; DB 15; Length 335;
Best Local Similarity 37.7%; Pred. No. 5.9e-36;
Matches 120; Conservative 56; Mismatches 112; Indels 30; Gaps 8;

QY 12 DPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLAIYSQIVVELHNS 71
Db 13 DKSTQKEQDEILLQPTTYIQIPGKQFRSELALAFNHNLLIPGEKLAQIGNIVQMLHNS 72

QY 72 LIIDDIEDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLVSOLTTEKPLYHNLITIF 131
Db 72 LIIDDIEDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLVSOLTTEKPLYHNLITIF 131
```

Db 73 LL---IEDNSILRRGYPVAHSIYGVASTINAANYALFLALEKVQQLDHPE-----VRKVY 124
QY 132 NEELINLRHQGLDIYWRDFLPBIIPTQEMYLNMWNKTGGLFRLLRLMEALSPSSHG 191
Db 125 TEQLIELHRGQGWELIWRDSP--TCPSESDYKLMVVRKTGGFLMLAIRLMQLFSSNKEDY 182
QY 192 HSLVFPINLLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSPPIVHALNFTTKGT 251
Db 183 SKLTA---ILGLYFQIRDDYCNLSLKEVSVNKSFADLTEGKFGFPVHAVTQK----- 234
QY 252 EQHNEILRILLRTSDKDIKLIQILEPDTNSLAYTKNFINQLVNMIKND-----NENKY 307
Db 235 -QDKQVLRKSQRTHDIEVKYCITILE-KLGSFOYTRKVLESLEDAEARSSEVSTLGSNPY 292
QY 308 LPDLASH-----SDTAT 319
Db 293 MDRLNKLKLSWKTSDSAS 310

Search completed: October 17, 2004, 09:18:43
Job time : 33.4343 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:24:39 ; Search time 89,8123 Seconds
(without alignments)
1338.060 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINDDPVSSQNE.....DTATNLHDELYIIDHLSEL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|-------------|
| 1 | 1720 | 100.0 | 335 | 5 | AAC22195 |
| 2 | 1720 | 100.0 | 335 | 5 | ABG68086 |
| 3 | 1720 | 100.0 | 335 | 7 | ADC49389 |
| 4 | 1720 | 100.0 | 335 | 8 | ADL14811 |
| 5 | 1720 | 100.0 | 335 | 8 | ADM98602 |
| 6 | 753 | 43.8 | 320 | 7 | ADF75081 |
| 7 | 612.5 | 35.6 | 343 | 6 | ADP96126 |
| 8 | 591.5 | 34.4 | 303 | 8 | ADM98614 |
| 9 | 583 | 33.9 | 300 | 3 | AAV68909 |
| 10 | 583 | 33.9 | 300 | 8 | ADJ10383 |
| 11 | 583 | 33.9 | 300 | 8 | ADM98626 |
| 12 | 577 | 33.5 | 304 | 3 | AAV58276 |
| 13 | 574 | 33.4 | 300 | 2 | AAW59753 |
| 14 | 574 | 33.4 | 300 | 2 | AAV27005 |
| 15 | 570 | 33.1 | 300 | 2 | AAV97565 |
| 16 | 545.5 | 31.7 | 338 | 4 | ABB60947 |
| 17 | 545.5 | 31.7 | 338 | 8 | ADM98657 |
| 18 | 545.5 | 31.7 | 338 | 8 | ADM98604 |
| 19 | 538 | 31.3 | 376 | 2 | AAW22498 |
| 20 | 513 | 29.8 | 335 | 8 | ADM98656 |
| 21 | 513 | 29.8 | 335 | 8 | ADM98603 |
| 22 | 437.5 | 25.4 | 320 | 4 | AAU30743 |
| 23 | 337 | 19.6 | 347 | 6 | ABU19165 |
| 24 | 274.5 | 16.0 | 364 | 6 | ABU48648 |
| 25 | 246.5 | 14.3 | 322 | 6 | ABU40267 |

| | | | | | |
|----|-------|------|-----|---|----------|
| 26 | 243.5 | 14.2 | 323 | 6 | ABU50285 |
| 27 | 242.5 | 14.1 | 307 | 6 | ABU40769 |
| 28 | 242.5 | 14.1 | 307 | 7 | ADF05623 |
| 29 | 239 | 13.9 | 325 | 7 | ADB74153 |
| 30 | 238.5 | 13.9 | 324 | 6 | ABU38162 |
| 31 | 236.5 | 13.8 | 330 | 2 | AAW12391 |
| 32 | 236 | 13.7 | 324 | 6 | ABP78926 |
| 33 | 236 | 13.7 | 324 | 6 | ABU37471 |
| 34 | 233 | 13.5 | 322 | 4 | AAU36438 |
| 35 | 233 | 13.5 | 322 | 6 | ABU38762 |
| 36 | 233 | 13.5 | 341 | 7 | ABO72702 |
| 37 | 231.5 | 13.5 | 330 | 2 | AAW12389 |
| 38 | 230.5 | 13.4 | 324 | 6 | ABM68417 |
| 39 | 229.5 | 13.3 | 330 | 2 | AAW12386 |
| 40 | 229.5 | 13.3 | 330 | 2 | AAW12390 |
| 41 | 228.5 | 13.3 | 322 | 6 | ABU41575 |
| 42 | 228.5 | 13.3 | 330 | 2 | AAV79969 |
| 43 | 228.5 | 13.3 | 330 | 2 | AAW52283 |
| 44 | 228.5 | 13.3 | 330 | 2 | AAW52282 |
| 45 | 228.5 | 13.3 | 330 | 2 | AAW62528 |

ALIGNMENTS

RESULT 1

AAO22195
ID AAO22195 standard; protein; 335 AA.

XX AC AAO22195;

XX DT 03-OCT-2002 (first entry)

XX DE Protein relating to the production of prenyl alcohol SEQ ID No 6.

XX KW Prenyl alcohol; mutated cell; squalene synthase gene; geometric isomer;
XX KW industrial synthesis; isoprenoid-terpenoid compound.

XX OS Saccharomyces cerevisiae.

XX FN WO200253747-A1.

XX PD 11-JUL-2002.

XX PF 20-DEC-2001; 2001WO-JF011215.

XX PR 28-DEC-2000; 2000JP-00401701.

XX PR 28-DEC-2000; 2000JP-00403067.

XX PR 18-SEP-2001; 2001JP-00282978.

XX (TOYT) TOYOTA JIDOSHA KK.

XX Ohto C, Obata S;

XX WPI; 2002-548086/58.

XX N-PSDB; AAL40786.

XX Production of prenyl alcohols by culturing translationally-active mutated cells with reduced squalene synthase gene to express less transcriptional product.

XX Claim 16; Page 146-148; 262pp; Japanese.

XX The invention relates to a method for producing prenyl alcohol comprising culturing mutated cells, having been mutated so as to reduce the amount of transcriptional product of squalene synthase gene transcriptional activity, and then collecting prenyl alcohol from the culture medium. The method is for the production of prenyl alcohols, which is for use in industrial synthesis of isoprenoid-terpenoid compounds particularly physiologically-active prenyl alcohol geometric isomers. This sequence represents a protein relating to the method for producing prenyl alcohol comprising culturing mutated cells

SQ Sequence 335 AA;

Query Match 100.0%; Score 1720; DB 5; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60
 DB 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60

QY 61 SOIVELLHNSLLIIDDIEDNAPLRGGQTTHLIFGVPTINTANTMYFRAMOLVSQLTAK 120
 DB 61 SOIVELLHNSLLIIDDIEDNAPLRGGQTTHLIFGVPTINTANTMYFRAMOLVSQLTAK 120

QY 121 EPLVHNLTIFNEELINLHGGGLDIYWRDFLPEIIPQEMVYNMVKTKGLFELTLRL 180
 DB 121 EPLVHNLTIFNEELINLHGGGLDIYWRDFLPEIIPQEMVYNMVKTKGLFELTLRL 180

QY 181 MEALSPSSHGHSLVPPFNILGIIYQIRDDYLNKDFQMSSEKGFADITEGKLSPIVH 240
 DB 181 MEALSPSSHGHSLVPPFNILGIIYQIRDDYLNKDFQMSSEKGFADITEGKLSPIVH 240

QY 241 ALNFTKTKGQTEQHNEILRLILRTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMVK 300
 DB 241 ALNFTKTKGQTEQHNEILRLILRTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMVK 300

QY 301 NDENKYLPLASHSDTATNLHDELLYIIDHLSL 335
 DB 301 NDENKYLPLASHSDTATNLHDELLYIIDHLSL 335

RESULT 2
 ABG68086
 ID ABG68086 standard; protein; 335 AA.
 AC ABG68086;
 XX
 XX
 XX 24-SEP-2002 (first entry)
 DE
 DE S. cerevisiae prenyl diphosphate synthase #2.
 XX
 XX Prenyl alcohol; prenyl diphosphate synthase; geranylgeraniol;
 KW hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; enzyme;
 KW isopentenyl diphosphate delta-isomerase; mevalonate kinase;
 KW mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.
 XX
 XX Saccharomyces cerevisiae.
 OS
 XX
 XX WO200253746-A1.
 XX
 XX 11-JUL-2002.
 PD
 XX
 XX 20-DEC-2001; 2001WO-JP011214.
 XX
 XX 28-DEC-2000; 2000JP-00403067.
 XX
 XX (TOYT) TOYOTA JIDOSHA KK.
 PA
 XX
 XX Ohto C, Obata S, Muramatsu M, Nishi K, Totsuka K;
 PI
 XX
 XX WPI; 2002-537944/57.
 DR
 XX
 XX Production of prenyl alcohols by culturing a transformant transferred
 PT with e.g. prenyl diphosphate synthase gene, for use in industrial
 PT synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds.
 XX
 XX Disclosure; Page 185-187; 335pp; Japanese.
 XX
 XX The invention relates to a process for producing a prenyl alcohol
 CC comprising: (A) construction of a recombinant by transferring an
 CC expression recombinant DNA or a DNA for genome integration into a host
 CC which contains prenyl diphosphate synthase gene or its variant; and (B)
 CC collecting product from the culture medium. Also described is: (1) a

method for producing prenyl alcohol in which the expression recombinant
 DNA contains: (a) a hydroxymethylglutaryl-CoA reductase gene or its
 variant; or (b) an isopentenyl diphosphate delta-isomerase gene; (2) a
 process for producing geranylgeraniol, comprising: (a) constructing a
 recombinant by transferring an expression recombinant DNA or a DNA for
 genome integration into a host which contains hydroxymethylglutaryl-CoA
 reductase gene or its variant; and (b) isolating the product; (3) a
 process for producing farnesylgeraniol in which the expression
 recombinant DNA also contains an isopentenyl diphosphate delta-isomerase
 gene, mevalonate CoA acetyltransferase gene, hydroxymethylglutaryl-CoA
 synthase gene, mevalonate kinase gene, or mevalonate diphosphate
 decarboxylase gene. The methods are used for the production of prenyl
 alcohols, particularly for use in industrial synthesis of isoprenoid-
 terpenoid compounds e.g. physiologically-active prenyl alcohols including
 geometric isomers. ABG68084-ABG68095 represent prenyl diphosphate
 synthase proteins used in the methods of the invention

SQ Sequence 335 AA;

Query Match 100.0%; Score 1720; DB 5; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60
 DB 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60

QY 61 SOIVELLHNSLLIIDDIEDNAPLRGGQTTHLIFGVPTINTANTMYFRAMOLVSQLTAK 120
 DB 61 SOIVELLHNSLLIIDDIEDNAPLRGGQTTHLIFGVPTINTANTMYFRAMOLVSQLTAK 120

QY 121 EPLVHNLTIFNEELINLHGGGLDIYWRDFLPEIIPQEMVYNMVKTKGLFELTLRL 180
 DB 121 EPLVHNLTIFNEELINLHGGGLDIYWRDFLPEIIPQEMVYNMVKTKGLFELTLRL 180

QY 181 MEALSPSSHGHSLVPPFNILGIIYQIRDDYLNKDFQMSSEKGFADITEGKLSPIVH 240
 DB 181 MEALSPSSHGHSLVPPFNILGIIYQIRDDYLNKDFQMSSEKGFADITEGKLSPIVH 240

QY 241 ALNFTKTKGQTEQHNEILRLILRTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMVK 300
 DB 241 ALNFTKTKGQTEQHNEILRLILRTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMVK 300

QY 301 NDENKYLPLASHSDTATNLHDELLYIIDHLSL 335
 DB 301 NDENKYLPLASHSDTATNLHDELLYIIDHLSL 335

RESULT 3
 ADC49389
 ID ADC49389 standard; protein; 335 AA.
 XX
 XX ADC49389;
 AC
 XX
 XX 18-DEC-2003 (first entry)
 DT
 XX
 XX Yeast BTS1 protein.
 DE
 XX
 XX Yeast; microbiological; prenyl alcohol; geranylgeraniol; farnesol;
 KW nerolidol; squalene synthetase inhibitor; BTS1.
 XX
 XX Saccharomyces cerevisiae.
 OS
 XX
 XX JP2002300896-A.
 PN
 XX
 XX 15-OCT-2002.
 PD
 XX
 XX 18-JAN-2002; 2002JP-00010528.
 XX
 XX 30-JAN-2001; 2001JP-00021547.
 XX
 XX (TOYT) TOYOTA JIDOSHA KK.
 PA
 XX

DR WPI; 2003-485780/46.
 DR N-PSDB; ADC49385.
 XX
 PT A process for preparation of prenyl alcohols by culture of their
 PT productive microorganisms in a medium with added squalene synthetase
 PT inhibitor.
 XX
 PS Disclosure; SEQ ID NO 12; 37pp; Japanese.
 XX
 CC The invention discloses a microbiological preparation of prenyl alcohols,
 CC particularly geranylgeraniol, farnesol and/or nerolidol, by culture of
 CC their productive microorganism spp. of Saccharomyces, Saccharomycopsis,
 CC Saccharomycodes, Schizosaccharomyces, Wickerhamia, Debaryomyces,
 CC Hanseniaspora, Pichia, Candida, Zygosaccharomyces, Ogataea, Kurashia,
 CC Komagataella, Yarrowia, Kluyveromyces, Trichosporon, Cryptococcus,
 CC Torulaspora, Bacillus, Staphylococcus, Micrococcus, Exiguobacterium,
 CC Gibberella or Mucor, or Ambrosiozyma, Cystofilobasidium, Metschnikowia,
 CC Trichosporiella, Xanthophyllomyces, Bullera, Fellomyces, Filobasidium,
 CC Holtermannia, Phaffia, Rhodotorula, Sporidiobolus, Sporobolomyces,
 CC Williopsis, Zygosascus, Leucosporidium, Myxozyma, Klöckeria,
 CC Brevibacterium, Alcaligenes or Haloferax in a medium with added squalene
 CC synthetase inhibitor to extra- and intra-cellular cumulate prenyl
 CC alcohols, followed by their collection. The sequence presented is the
 XX yeast HWG1 cDNA.
 XX
 SQ Sequence 335 AA;
 Query Match 100.0%; Score 1720; DB 7; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPKGNFRNLIVQINRVNMLPKDQLATV 60
 Db 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPKGNFRNLIVQINRVNMLPKDQLATV 60
 QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120
 Db 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120
 QY 121 EPLVHNLITIFNEELINLHRRGGGLDYWRDFELPEIIPTEQMYLNVMNKTGGLFRLTLRL 180
 Db 121 EPLVHNLITIFNEELINLHRRGGGLDYWRDFELPEIIPTEQMYLNVMNKTGGLFRLTLRL 180
 QY 181 MEALSPSSHHGSLVFFINLLGIYQIRDDYLNLDKDFQMSSEKGFPAEDITGKLSFPIVH 240
 Db 181 MEALSPSSHHGSLVFFINLLGIYQIRDDYLNLDKDFQMSSEKGFPAEDITGKLSFPIVH 240
 QY 241 ALNFTKTKGQTEQHNELIRILLRTSDKDIKLLIQILEFDTNSLAYTKNFINQLVNMVIX 300
 Db 241 ALNFTKTKGQTEQHNELIRILLRTSDKDIKLLIQILEFDTNSLAYTKNFINQLVNMVIX 300
 QY 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSL 335
 Db 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSL 335
 RESULT 4
 ADL14811
 ID ADL14811 standard; protein; 335 AA.
 XX
 AC ADL14811;
 XX
 DT 17-JUN-2004 (first entry)
 XX
 DE Yeast geranylgeranyl diphosphate synthase.
 XX
 KW Yeast; enzyme; Isopentenyl diphosphate biosynthetic pathway; IPP;
 KW prenyl alcohol; squalene synthase.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN US2004063182-A1.
 XX

PD 01-APR-2004.
 XX
 PF 18-JUN-2003; 2003US-00450941.
 XX
 PR 28-DEC-2000; 2000JP-00401701.
 PR 28-DEC-2000; 2000JP-00406067.
 PR 18-SEP-2001; 2001JP-00282978.
 PR 20-DEC-2001; 2001WO-JF011215.
 XX
 XX (OHTO/) OHTO C.
 PA (OBAT/) OBATA S.
 PA
 XX Ohto C, Ohta S;
 XX WPI; 2004-294401/27.
 DR N-PSDB; ADL14810.
 PT
 PT Producing prenyl alcohols by culturing a mutant cell in which the amount
 PT of squalene synthase gene transcript having translational activity is
 PT reduced.
 XX
 PS Claim 16; SEQ ID NO 6; 139pp; English.
 XX
 CC The invention relates to producing a prenyl alcohol comprising culturing
 CC a mutant cell that has been mutated so that an amount of squalene
 CC synthase gene transcript having translational activity can be reduced,
 CC and recovering the prenyl alcohol from the resultant culture.
 CC Alternatively, producing a prenyl alcohol also comprises culturing a
 CC mutant cell in which the transcription promoter region of its squalene
 CC synthase gene has been replaced with a transcription repression-type
 CC promoter under transcription repression conditions, reducing an amount of
 CC squalene synthase gene transcript having translational activity, and
 CC recovering the prenyl alcohol from the resultant culture. Also included
 CC are a method of producing a prenyl alcohol (comprising preparing a
 CC recombinant by introducing a recombinant DNA for expression or a DNA for
 CC genomic integration each comprising an IPP (isoprenyl diphosphate)
 CC biosynthetic pathway-related enzyme gene into any of the mutant cells
 CC cited above), a mutant cell that has been mutated so that an amount of
 CC squalene synthase gene transcript having translational activity can be
 CC reduced) and a mutant cell in which the transcription promoter region of
 CC its squalene synthase gene has been replaced with a transcription
 CC repression-type promoter so that an amount of transcript from its
 CC squalene synthase gene having translational activity can be reduced. The
 CC transcription repression-type promoter in producing a prenyl alcohol is
 CC GAL1 promoter. The IPP biosynthetic pathway-related enzyme gene is any
 CC one selected from farnesyl diphosphate synthase gene, geranylgeranyl
 CC diphosphate synthase gene, hydroxymethylglutaryl-CoA reductase gene,
 CC isopentenyl diphosphate A-isomerase gene, neovalonate kinase gene, acetyl-
 CC CoA acetyltransferase gene, hydroxymethylglutaryl-CoA synthase gene,
 CC phosphomevalonate kinase gene and diphosphomevalonate decarboxylase gene,
 CC or a fusion gene composed of a gene selected from the above genes, or a
 CC gene obtained by introducing an addition, substitution or insertion
 CC mutation into any one of the above genes so that the polypeptide encoded
 CC by the resultant gene contains an endoplasmic reticulum signal. The
 CC methods and compositions of the present invention are useful for
 CC producing prenyl alcohols by reducing the amount of squalene synthase
 CC gene transcript having translational activity. The present sequence
 CC represents an IPP (isoprenyl diphosphate) biosynthetic pathway-related
 CC enzyme.
 XX
 SQ Sequence 335 AA;
 Query Match 100.0%; Score 1720; DB 8; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPKGNFRNLIVQINRVNMLPKDQLAIV 60
 Db 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPKGNFRNLIVQINRVNMLPKDQLAIV 60
 QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120
 Db 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120

QY 121 EPLYHNLITIFNEELINLHRCGLDIYWRDLPFLPIIPTQEMYLNVMNKTGGLFRLTLRL 180
Db 121 EPLYHNLITIFNEELINLHRCGLDIYWRDLPFLPIIPTQEMYLNVMNKTGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
QY 241 ALNFTTKGTQOEHNEILRLILLRTSDKDIKLIQILEFDNLSLAYTKNFNLQVNMNIK 300
Db 241 ALNFTTKGTQOEHNEILRLILLRTSDKDIKLIQILEFDNLSLAYTKNFNLQVNMNIK 300
QY 301 NDNENKYLPLDASHSDTATNLHDELIIIDHSEL 335
Db 301 NDNENKYLPLDASHSDTATNLHDELIIIDHSEL 335

RESULT 5

ADM98602
ID ADM98602 standard; protein; 335 AA.
XX
AC
XX ADM98602;
XX
DT 01-JUL-2004 (first entry)
XX Geranylgeranyl pyrophosphate synthase polypeptide #1.
DE
KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
KW diterpene synthase; defence toxin; volatile defensive signal;
KW pollinator attractant; photoprotectant; enzyme.
XX
OS Saccharomyces cerevisiae.
XX
PN US2004072323-A1.
XX
PD 15-APR-2004.
XX
PF 07-JAN-2002; 2002US-00041018.
XX
PR 05-JAN-2001; 2001US-0259880P.
XX
PA (MATS/) MATSUDA S P T.
PA (HART/) HART E A.
XX
PI Matsuda SPT, Hart EA;
XX
XX WPI; 2004-373921/35.
XX
XX New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
PS Claim 30; SEQ ID NO 22; 38pp; English.
XX

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing

CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a geranylgeranyl pyrophosphate synthase polypeptide
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification but was obtained in electronic format
CC from USPTO at seqdata.uspto.gov/sequence.html.

XX
SQ Sequence 335 AA;

Query Match 100.0%; Score 1720; DB 8; Length 335;
Best Local Similarity 100.0%; Pred. No. 4.6e-154;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAKIDELINNDPVWSSQNESLISKPYNHILLKPGKNFRLNLIQVNRVWNLPKQQLAIV 60
Db 1 MEAKIDELINNDPVWSSQNESLISKPYNHILLKPGKNFRLNLIQVNRVWNLPKQQLAIV 60
QY 61 SQIVELHNSLLIIDDIEDNAPLRGQTTSHLIFGVPTINTANYMYFAMQVLSQLTTK 120
Db 61 SQIVELHNSLLIIDDIEDNAPLRGQTTSHLIFGVPTINTANYMYFAMQVLSQLTTK 120
QY 121 EPLYHNLITIFNEELINLHRCGLDIYWRDLPFLPIIPTQEMYLNVMNKTGGLFRLTLRL 180
Db 121 EPLYHNLITIFNEELINLHRCGLDIYWRDLPFLPIIPTQEMYLNVMNKTGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
QY 241 ALNFTTKGTQOEHNEILRLILLRTSDKDIKLIQILEFDNLSLAYTKNFNLQVNMNIK 300
Db 241 ALNFTTKGTQOEHNEILRLILLRTSDKDIKLIQILEFDNLSLAYTKNFNLQVNMNIK 300
QY 301 NDNENKYLPLDASHSDTATNLHDELIIIDHSEL 335
Db 301 NDNENKYLPLDASHSDTATNLHDELIIIDHSEL 335

RESULT 6

ADF75081
ID ADF75081 standard; protein; 320 AA.
XX
AC ADF75081;
XX
DT 26-FEB-2004 (first entry)
XX
DE A_gossypii protein essential for fungal growth & development SeqID 22.
XX
KW fungal growth; fungicide; screening assay; fungicidal; crop.
XX
OS Bremothecium gossypii.
XX
PN WO2003072706-A2.
XX
PD 04-SEP-2003.
XX
PF 11-FEB-2003; 2003WO-US0004154.
XX
PR 22-FEB-2002; 2002US-0359557P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Gaffney TD, Flavler A, Dietrich F, Philippsen P, Kirksey CMM;
PI Gates K, Gardner J;
XX
XX WPI; 2003-721756/68.
XX
XX N-PSDB; ADF75080.

XX Identifying a fungicidal compound comprising combining a polypeptide with
XX the compound to be tested with the ability to bind, or to inhibit the
XX activity of the polypeptide under conditions conducive to binding or
XX inhibition.

15-APR-2004.
07-JAN-2002; 2002US-00041018.
05-JAN-2001; 2001US-0259880P.
(MATS/) MATSUDA S P T.
(HART/) HART E A.
Matsuda SPT, Hart EA;
WPI; 2004-373921/35.
New unicellular organisms comprising exogenous nucleic acids encoding a geranylgeranyl pyrophosphate and a diterpene synthase, useful for producing diterpenes and diterpene precursors.
Claim 30; SEQ ID NO 34; 38pp; English.
The invention relates to a unicellular organism for producing a diterpene or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polyaromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents a geranylgeranyl pyrophosphate synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.
Sequence 303 AA;
Query Match 34.4%; Score 591.5; DB 8; Length 303;
Best Local Similarity 41.1%; Pred. No. 3e-47;
Matches 113; Conservative 67; Mismatches 78; Indels 17; Gaps 5;
QY 20 ESLISKPYNHILKPGKFNELNLIQVIRVMNLPKDQALVSOIVVELHNSLLIDIED 79
DB 14 EDILFETYLISQPKDHRKALISAPDLMLHVPKDVLCVINKLIGHNASHMIDVDQ 73
QY 80 NAPLRGQTTSHLIFGVPESTINTANYMYFRAMQLVSOITTKPELYHNLIIFNEELINLH 139
DB 74 DSDLRGVPVAHHYGVPEQINTANYVIFLALQEVKMLNIP-----SMQVCTEELINLH 128
QY 140 RGGLDIYWRDFELPIIPTQEMYNLMVNTKGGIFRLTLRLMALSPSSHGSLVFPFN 199
DB 129 RGQGLIYWRDSL--TCPTTEEYIDVMNNTSGLLRLAVLMOAASED---IDYPLVN 183
QY 200 LLIIGIYQIRDDYLNKDFQMSSEKGAEDITEGKLSFPIVHALNFTKQGTQHNEILR 259
DB 184 IIGIHQVRDDYVNLQSTYNNKGFCEDETEGKFSFPIIHAIR-----KDPNSRQLN 237
QY 260 ILLRLTSDDIKLKIQLIEFDNLSLAYTKNFNINQ 294
DB 238 IISQKPTSEVKKYALEVIR-KAGSFYEVREFLQ 271
RESULT 9
AAV68909

AAV68909 standard; protein; 300 AA.
AAV68909;
16-MAY-2000 (first entry)
A geranylgeranyl pyrophosphate synthetase (hGGPPS).
Human; geranylgeranyl pyrophosphate synthetase; hGGPPS; chromosome 1; 1q42-1q43 locus; prostate cancer; hGGPPS; biallelic marker; mevalonic biosynthetic pathway.
Homo sapiens.
WO200005382-A2.
03-FEB-2000.
23-JUL-1999; 99WO-IB001353.
23-JUL-1998; 98US-0093940P.
(GEST) GENSET.
Bougueleret L;
WPI; 2000-182704/16.
N-PSDB; AAZ60888, AAZ60889, AAZ60890.
New isolated human geranyl-geranyl pyrophosphate synthetase nucleic acids, used to develop agents for the diagnosis of, e.g. pathologies related to a defect in the mevalonic biosynthetic pathway.
Claim 15; Page 84-85; 88pp; English.
The present sequence represents a human geranylgeranyl pyrophosphate synthetase (hGGPPS). Two differently spliced mRNAs exist for this gene. The first spliced mRNA is derived from a cDNA (AAZ60888) which comprises exons 1, 2, 3 and 4. The second mRNA is derived from a cDNA (AAZ60889) which comprises 1b1s, 2, 3, and 4. The hGGPPS gene is located on chromosome 1, at the 1q42-1q43 locus. This chromosome 1 locus has been shown to carry a predisposing gene for prostate cancer. The nucleic acids encoding hGGPPS can be used for screening for agents which modulate the expression of the hGGPPS gene. Such agents can be used in therapeutic applications. The biallelic markers associated with the hGGPPS gene can be used for the diagnosis of diseases related to an alteration in the regulatory or coding regions of hGGPPS, such as pathologies related to a defect in the mevalonic biosynthetic pathway. The products can also be used for detection, diagnosis and drug screening
Sequence 300 AA;
Query Match 33.9%; Score 583; DB 3; Length 300;
Best Local Similarity 40.5%; Pred. No. 1.9e-46;
Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;
QY 20 ESLISKPYNHILKPGKFNELNLIQVIRVMNLPKDQALVSOIVVELHNSLLIDIED 79
DB 9 QRILLEPYKLLQPKGVKRTKLSQAFNEHWLKVPEKQLQIIIEVTEMLHNASLLIDIED 68
QY 80 NAPLRGQTTSHLIFGVPESTINTANYMYFRAMQLVSOITTKPELYHNLIIFNEELINLH 139
DB 69 NSKLRRGFPVAHHYGVPEQINTANYVIFLALQEVKMLNIP-----SMQVCTEELINLH 123
QY 140 RGGLDIYWRDFELPIIPTQEMYNLMVNTKGGIFRLTLRLMALSPSSHGSLVFPFN 199
DB 124 QGQGLIYWRD--NYTCPTTEEYKAMVLQTKGFLGVLGMLQIF---SDYKEDLKLPLN 178
QY 200 LLIIGIYQIRDDYLNKDFQMSSEKGAEDITEGKLSFPIVHALNFTKQGTQHNEILR 259
DB 179 TLGLFFQIRDDYVNLQSTYNNKGFCEDETEGKFSFPIIHAIR-----WSRPESTOVQN 232
QY 260 ILLRLTSDDIKLKIQLIEFDNLSLAYTKNFNINQVNMKNENKYLPLASHSDTAT 319

Db 233 ILRQNTENIDIKKVCVHYLE-DVGSFEYTRNLTKEK-----EAK-----AYKQIDAR 278

QY 320 N1HDELLV1IDHSEL 335

Db 279 GGNPELV1ALVKHLSKM 294

RESULT 10

ADJ10383

ID ADJ10383 standard; protein; 300 AA.

AC ADJ10383;

XX 17-JUN-2004 (first entry)

DT Human geranylgeranyl diphosphate synthase 1 protein.

DE human; geranylgeranyl diphosphate synthase 1; GGPS1;

XX geranylgeranyl pyrophosphate synthetase; GGPPS; ggppsase;

KW geranyltransferase; embryonic development; cell differentiation;

KW apoptosis; developmental; hyperproliferative disorder; cancer;

KW cytostatic.

XX

OS Homo sapiens.

XX

XX US2004005570-A1.

XX

XX 08-JAN-2004.

XX

XX 02-JUL-2002; 2002US-00189268.

XX

XX 02-JUL-2002; 2002US-00189268.

XX (ISIS-) ISIS PHARM INC.

PA Dean NM, Bennett CF, Dobie KW;

PI WPI; 2004-201146/19.

XX N-PSDB; ADJ10255, ADJ10255.

XX

XX New antisense oligonucleotides for modulating geranylgeranyl diphosphate

PT synthase 1 expression, useful for diagnosing, preventing or treating

PT conditions associated with the protein, e.g. cancer.

XX

XX Disclosure; Page 31-32; 76pp; English.

XX

XX This invention relates to a novel antisense compounds that modulate the

CC expression of human geranylgeranyl diphosphate synthase 1 (also known as

CC GGPS1, geranylgeranyl pyrophosphate synthetase, GGPPS, ggppsase and

CC geranyltransferase) and located on chromosome 1p43. Specifically, it

CC refers to compositions useful for inhibiting the expression of GGPS1,

CC which normally participates in embryonic development, cell

CC differentiation and stimulates apoptosis via caspase-3 activation. The

CC present invention describes antisense oligonucleotides that comprise at

CC least one modified sugar moiety, a 2'-O-methoxyethyl (2', MOE) and at

CC least one modified nucleobase, a 5-methylcytosine. Accordingly, these

CC compounds are useful for treating a disease or condition associated with

CC GGPS1 such as a developmental or hyperproliferative disorder (e.g.

CC cancer) that arise as a result of aberrant apoptosis. As such, these

CC compositions exhibit cytostatic activity and are useful for diagnostics,

CC prophylaxis, research reagents and various kits. This polypeptide

CC sequence is the geranylgeranyl diphosphate synthase 1 protein of the

CC invention. NOTE: This sequence is given as an embedded protein in the

CC sequence listing and is not further referred to in the specification.

XX

XX Sequence 300 AA;

XX

Query Match 33.9%; Score 583; DB 8; Length 300;

Best Local Similarity 40.5%; Pred. No. 1.9e-46;

Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;

QY 20 ESLISKPNHLLKPGKFNRLIVQINRVNMLPKDQIAIVSQIVELLHNSLLIDIED 79

Db 9 QRILLEPKYLQLPCKQVRKLSQAFNHWLKPEDKQLQIIIEVTMLHNASLIDIED 68

QY 80 NAPLRGQTTSHLIFGVPSVTINTANYMYFRAMQLVSLTTEKPLYHNLIITINEELINLH 139

Db 69 NSKLRRGFPVAHSIYIGIPSVINSANYVFLGLEKV--LTDHP---DAVKLFTROLLELH 123

QY 140 RQGLDIYWRDPLPEIIPTOEYMLNWNKVTGGLFRLRLRLMEALSPSSHGHSHSVPTIN 199

Db 124 QGQGLDIYWRD--NYTCTPEEYKAMVLQKTGGLFGLAVGLMQLF---SDYKEDLKPLIN 178

QY 200 LLGIYQIRDDYLNKQFQMSSEKGAEDITEGKLSFPVIVHALNFTKTKGQTEQNEILR 259

Db 179 TLGLFFQIRDDYANLHSEYSENKSCFCELTGKPSFTTTHAI-----WSRPESTQVON 232

QY 260 ILLRTSKDKIKLKIQLIEPDTNSLAYTKNFINQLVNMKNKNENKYLPLDASHSDTAT 319

Db 233 ILRQNTENIDIKKVCVHYLE-DVGSFEYTRNLTKEK-----EAK-----AYKQIDAR 278

QY 320 N1HDELLV1IDHSEL 335

Db 279 GGNPELV1ALVKHLSKM 294

RESULT 11

ADM98626

ID ADM98626 standard; protein; 300 AA.

XX

XX ADM98626;

XX

XX 01-JUL-2004 (first entry)

XX

XX Geranylgeranyl pyrophosphate synthase polypeptide #25.

XX

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

KW diterpene synthase; defence toxin; volatile defensive signal;

KW pollinator attractant; photoprotectant; enzyme.

XX

XX Homo sapiens.

XX

XX US2004072323-A1.

XX

XX 15-APR-2004.

XX

XX 07-JAN-2002; 2002US-00041018.

XX

XX 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX

XX Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

XX

XX New unicellular organisms comprising exogenous nucleic acids encoding a

PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for

PT producing diterpenes and diterpene precursors.

XX

XX Claim 30; SEQ ID NO 46; 38pp; English.

XX

XX The invention relates to a unicellular organism for producing a diterpene

CC or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a

CC promoter operable in the organism, and an exogenous nucleic acid sequence

CC encoding a diterpene synthase under the control of a promoter operable in

CC the organism. The invention also relates to methods of producing a

CC diterpene or diterpene precursor and a method of isolating a diterpene

CC synthase comprising growing several cells in the presence of a

CC polyaromatic resin to make a cell/resin mixture, where at least one of

CC the cells further comprises at least one isolated and purified nucleic

CC acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under

CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents a geranylgeranyl pyrophosphate synthase polypeptide
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification but was obtained in electronic format
 CC from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 300 AA;

Query Match 33.9%; Score 583; DB 8; Length 300;
 Best Local Similarity 40.5%; Pred. No. 1.9e-46;
 Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;

QY 20 ESLISKPNHLLKPKGNFRLMLIVQINRVMLPKDOLAIVSQIVELHNSLLDDIED 79
 DB 9 QRLLEPKYLLQLPKQVTKLSQAFNHLKVPEDKQLIIEVTMLHNSLLDDIED 68
 QY 80 NAPLRGQVTSHLIFGVPTINTANYMYPRAMQVLSQVTKKPLVHNLITIFNEELINLH 139
 DB 69 NSKLRRGFPVAHSYGIPSVINSANVYFLGLEKV--LTLDHP---DAVKLFRQLLELH 123
 QY 140 RQGLDIYWRDFELPIPTQEMVLMNMVMTKGLFRLTLRLMEALSPSHHSHLVFFIN 199
 DB 124 QOGLDIYWRD--NYTCPTTEERYKAMVLQKTGFLGLAVGLMQLF---SDYKEDLKPLIN 178
 QY 200 ILGITVQIRDDVNLKDFQMSSEKGFADITSGKSPFVIVHALNFTKTKQTEQHNELIR 259
 DB 179 TLGLFFQIRDDVNLKDFQMSSEKGFADITSGKSPFVIVHALNFTKTKQTEQHNELIR 232
 QY 260 ILLTSDTDKDKIKLITQILEFFDTNSLAYTKNFINQIVNMKNNDENKYLPLDLASHSDTAT 319
 DB 233 ILRQRTENIDIKKYCVHYLE-DVGSFEYTRNTLKEI-----EAK-----AYKQIDAR 278
 QY 320 NLHDELLYIDHLSL 335
 DB 279 GGNPELVALVKGHL SKM 294

RESULT 12
 ID AAB58276
 XX AAB58276 standard; protein; 304 AA.
 AC AAB58276;
 XX 14-MAR-2001 (first entry)
 DT Lung cancer associated polypeptide sequence SEQ ID 614.
 DE Human; lung cancer associated protein; neuroprotective; cytostatic;
 XX cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antiinfective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disease; reproductive;
 KW proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.
 OS
 XX WO200055180-A2.
 PN
 XX 21-SEP-2000.
 PD
 XX 08-MAR-2000; 2000WO-US005918.
 PF
 XX 12-MAR-1999; 99US-0124270P.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 PA
 XX

PI Ruben SM;
 XX WPI: 2000-587514/55.
 DR N-PSDB; AAF18152.
 XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer.
 XX Claim 11; Page 1108-1109; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotropic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the protein
 CC or polynucleotide sequences. The lung cancer associated polynucleotide
 CC sequences may be used for detection of lung cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The proteins may be used to treat disorders such as
 CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
 CC cardiovascular, renal, and proliferative disorders. The proteins may also
 CC be used in the treatment of wounds and infectious diseases.
 CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
 CC used in the course of the invention for the identification and
 CC characterisation of the polynucleotide and protein sequences

XX Sequence 304 AA;

Query Match 33.5%; Score 577; DB 3; Length 304;
 Best Local Similarity 39.6%; Pred. No. 7e-46;
 Matches 129; Conservative 58; Mismatches 107; Indels 32; Gaps 9;

QY 12 DPWSSQN--ESLISKPNHLLKPKGNFRLMLIVQINRVMLPKDOLAIVSQIVELHNL 69
 DB 3 NPMEKTQETVQRIILEPKYLLQLPKQVTKLSQAFNHLKVPEDKQLIIEVTMLHNL 62
 QY 70 SSLIIDDIEDNAPLRGQVTSHLIFGVPTINTANYMYPRAMQVLSQVTKKPLVHNLIT 129
 DB 63 ASLLIIDDIEDNAPLRGQVTSHLIFGVPTINTANYMYPRAMQVLSQVTKKPLVHNLIT 117
 QY 130 IFNEELINLHRRGQGLDIYWRDFELPIPTQEMVLMNMVMTKGLFRLTLRLMEALSPSSH 189
 DB 118 LFRQLLELHRRGQGLDIYWRD--NYTCPTTEERYKAMVLQKTGFLGLAVGLMQLF---SD 172
 QY 190 HGSHLVFFINLIGIYQIRDDVNLKDFQMSSEKGFADITSGKSPFVIVHALNFTKTKG 249
 DB 173 YKEDLKPLINLITGLFFQIRDDVNLKDFQMSSEKGFADITSGKSPFVIVHALNFTKTKG 226
 QY 250 QTEQHNELIRILLTSDTDKDKIKLITQILEFFDTNSLAYTKNFINQIVNMKNNDENKYL 309
 DB 227 SRXESTQVQNLIRQRTENIDIKKYCVHYLE-DVGSFEYTRNTLKEI-----EAK--- 274
 QY 310 DLASHSDTATNLHDELLYIDHLSL 335
 DB 275 --AYKQIDARGGNPELVALVKGHL SKM 298

RESULT 13
 ID AAW59753
 XX AAW59753 standard; protein; 300 AA.
 AC AAW59753;
 XX 12-OCT-1998 (first entry)
 DT Human geranyl pyrophosphate synthase.
 DE Human; geranyl pyrophosphate synthase; hGGPS; antagonist;
 XX agonist; ss.
 KW
 XX

```

OS Homo sapiens.
XX US5786193-A.
XX 28-JUL-1998.
XX
XX 06-JUN-1995; 95US-00469665.
XX
XX 06-JUN-1995; 95US-00469665.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Greene JM, Kirkness EF;
XX WPI; 1998-436529/37.
XX N-PSDB; AAV41555.
XX
XX DNA encoding human geranylgeranyl pyrophosphate synthetase polypeptide -
XX useful for producing recombinant polypeptide.
XX
XX Claim 1; Fig 1; 22pp; English.
XX
XX This is the amino acid sequence of the human geranyl geranyl
XX pyrophosphate synthase (hGPPS) used in the method of the invention. The
XX proteins are used in the production of recombinant hGPPS, which can be
XX used to screen for hGPPS antagonists or agonists
XX
XX Sequence 300 AA;
XX
XX Query Match 33.4%; Score 574; DB 2; Length 300;
XX Best Local Similarity 40.2%; Pred. No. 1.3e-45;
XX Matches 127; Conservative 55; Mismatches 104; Indels 30; Gaps 8;
XX
XX QY 20 ESLISKPNVHLLKPGKFNRLNLIQVNRVNNLPKDQIAIVSQIVELHNSLLIDIED 79
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 9 QRIILEPKYLLQLPGKQVTKLSQAFNHLKVPEDKQLIIRVTEMLHNASLLIDIED 68
XX
XX QY 80 NAPIRRGOTTSHLIFGVPTINTANYMYFRAMQLVSLTTKEPLVHNLITIFNEELINLH 139
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 69 NSKLRRGFPVAHSYIGIPSVINSANYVYFLGLEKV--LTLDDHP---DAVKLFTQLLELH 123
XX
XX QY 140 RGQGLDIYWRDFELPIIPTQEMVNNVNNKTKGGLFRITLRLMEALSPSSHHGSLVFPFN 199
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 124 QGQGLDIYWRD--NYTCPTTEEKYKAMVLOKTGGLFGLAVGLMQLF---SDYKEDLKPLN 178
XX
XX QY 200 LLGIYQIRDDYLNLDKDFOMSKSEKFAEDITEGKLSFPIVHNLFTKGTQTEQHNEILR 259
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 179 TLGLFFQIRDDYANLHSEKSEKSGEDITEGKFSFPTTHAI-----WSRSESTQVQN 232
XX
XX QY 260 ILLRTSDKDIKLLKLIQILEFDTNLSLAYTKNFNQLVNNMKNDNENKYLPLDASHSDTAT 319
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 233 ILRQRTENIDIKKYCVHYLB-DVSGSEYTRNTLKL-----EAK-----AYKQIDAR 278
XX
XX QY 320 NLHDELLIYIDHLSSEL 335
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 279 GGNPELVALVKHLSKM 294
XX
XX RESULT 14
XX ID AAY27005 standard; protein; 300 AA.
XX AC AAY27005;
XX XX
XX 20-SEP-1999 (first entry)
XX
XX Human geranylgeranyl pyrophosphate synthetase (hGPPS) polypeptide.
XX
XX Geranylgeranyl pyrophosphate synthetase; hGPPS; cell morphology; human;
XX apoptosis; neoplasia; tumour; cancer; viral proliferation; gene therapy;
XX human choroidermina.
XX
XX Homo sapiens.

```

```

XX US5928924-A.
XX
XX 27-JUL-1999.
XX
XX 11-MAR-1998; 98US-00038596.
XX
XX 11-JAN-1995; 95WO-US000421.
XX 06-JUN-1995; 95US-00469665.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Kirkness EF, Greene JM;
XX WPI; 1999-429504/36.
XX N-PSDB; AAX86773.
XX
XX Geranylgeranyl pyrophosphate synthetase peptides useful for regulating
XX apoptosis.
XX
XX Claim 1; Fig 1A-D; 23pp; English.
XX
XX This represents a human geranylgeranyl pyrophosphate synthetase (hGPPS)
XX polypeptide. The hGPPS polynucleotides and polypeptides may be used to
XX control the morphology of cells, suppress apoptosis, screen for hGPPS
XX agonists and antagonists (which may in turn be used to treat neoplasia,
XX e.g. tumours and cancer cell growth, and prevent viral proliferation),
XX gene therapy, and diagnose a disease or susceptibility to diseases
XX related to a mutation in hGPPS (e.g. human choroidermina)
XX
XX Sequence 300 AA;
XX
XX Query Match 33.4%; Score 574; DB 2; Length 300;
XX Best Local Similarity 40.2%; Pred. No. 1.3e-45;
XX Matches 127; Conservative 55; Mismatches 104; Indels 30; Gaps 8;
XX
XX QY 20 ESLISKPNVHLLKPGKFNRLNLIQVNRVNNLPKDQIAIVSQIVELHNSLLIDIED 79
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 9 QRIILEPKYLLQLPGKQVTKLSQAFNHLKVPEDKQLIIRVTEMLHNASLLIDIED 68
XX
XX QY 80 NAPIRRGOTTSHLIFGVPTINTANYMYFRAMQLVSLTTKEPLVHNLITIFNEELINLH 139
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 69 NSKLRRGFPVAHSYIGIPSVINSANYVYFLGLEKV--LTLDDHP---DAVKLFTQLLELH 123
XX
XX QY 140 RGQGLDIYWRDFELPIIPTQEMVNNVNNKTKGGLFRITLRLMEALSPSSHHGSLVFPFN 199
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 124 QGQGLDIYWRD--NYTCPTTEEKYKAMVLOKTGGLFGLAVGLMQLF---SDYKEDLKPLN 178
XX
XX QY 200 LLGIYQIRDDYLNLDKDFOMSKSEKFAEDITEGKLSFPIVHNLFTKGTQTEQHNEILR 259
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 179 TLGLFFQIRDDYANLHSEKSEKSGEDITEGKFSFPTTHAI-----WSRSESTQVQN 232
XX
XX QY 260 ILLRTSDKDIKLLKLIQILEFDTNLSLAYTKNFNQLVNNMKNDNENKYLPLDASHSDTAT 319
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 233 ILRQRTENIDIKKYCVHYLB-DVSGSEYTRNTLKL-----EAK-----AYKQIDAR 278
XX
XX QY 320 NLHDELLIYIDHLSSEL 335
XX Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 279 GGNPELVALVKHLSKM 294
XX
XX RESULT 15
XX ID AAR97565 standard; protein; 300 AA.
XX AC AAR97565;
XX XX
XX 17-FEB-1997 (first entry)
XX
XX Human geranylgeranyl pyrophosphate synthetase (GPPS).
XX
XX Geranylgeranyl pyrophosphate synthetase; cholesterol metabolism; steroid;
XX post-translational modification; gene therapy; rhadomyolysis;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:06:49 ; Search time 28.4406 Seconds
(without alignments)
781.156 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINNDPVWSSQNE.....DTATNLHDELVIIDHLSL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|----------------------|
| 1 | 1720 | 100.0 | 335 | 2 | US-08-761-344-2 |
| 2 | 574 | 33.4 | 300 | 1 | US-08-469-665-2 |
| 3 | 574 | 33.4 | 300 | 2 | US-09-038-596-2 |
| 4 | 574 | 33.4 | 300 | 5 | PCT-US95-00421-2 |
| 5 | 538 | 31.3 | 376 | 3 | US-09-091-725-15 |
| 6 | 242.5 | 14.1 | 307 | 4 | US-09-543-681A-5908 |
| 7 | 233 | 13.5 | 341 | 4 | US-09-252-991A-21448 |
| 8 | 232 | 13.5 | 165 | 4 | US-09-248-796A-17985 |
| 9 | 228.5 | 13.3 | 330 | 1 | US-08-410-167A-4 |
| 10 | 228.5 | 13.3 | 330 | 2 | US-08-898-560-1 |
| 11 | 227 | 13.2 | 342 | 4 | US-09-489-039A-13653 |
| 12 | 225.5 | 13.1 | 325 | 3 | US-09-217-609A-2 |
| 13 | 225.5 | 13.1 | 325 | 3 | US-08-873-235B-2 |
| 14 | 225.5 | 13.1 | 330 | 3 | US-09-101-126-1 |
| 15 | 221.5 | 12.9 | 340 | 4 | US-09-543-681A-5778 |
| 16 | 210.5 | 12.2 | 333 | 3 | US-09-025-819-29 |
| 17 | 210.5 | 12.2 | 333 | 4 | US-09-808-126-29 |
| 18 | 210.5 | 12.2 | 333 | 4 | US-09-803-951-29 |
| 19 | 206.5 | 12.0 | 162 | 4 | US-09-248-796A-17986 |
| 20 | 204 | 11.9 | 353 | 4 | US-09-328-352-4226 |
| 21 | 189 | 11.0 | 335 | 4 | US-09-673-018-2 |
| 22 | 188 | 10.9 | 380 | 4 | US-09-107-532A-4570 |
| 23 | 186.5 | 10.8 | 325 | 3 | US-09-009-895-2 |
| 24 | 174.5 | 10.1 | 319 | 4 | US-09-710-279-1040 |
| 25 | 174.5 | 10.1 | 357 | 3 | US-09-134-001C-3250 |
| 26 | 171.5 | 10.0 | 237 | 4 | US-09-934-903-14 |
| 27 | 171.5 | 10.0 | 297 | 4 | US-09-934-868-72 |

28 166.5 9.7 320 3 US-09-217-609A-6 Sequence 6, Appli
29 166.5 9.7 320 3 US-08-873-235B-6 Sequence 6, Appli
30 161.5 9.4 325 4 US-09-602-787A-260 Sequence 260, App
31 161.5 9.4 350 4 US-09-602-787A-258 Sequence 258, App
32 157.5 9.2 302 3 US-09-134-001C-3171 Sequence 3171, App
33 156.5 9.1 306 4 US-09-134-000C-4155 Sequence 4155, Ap
34 156 9.1 299 4 US-09-489-039A-12733 Sequence 12733, A
35 155 9.0 321 4 US-09-549-848B-12 Sequence 12, Appli
36 148.5 8.6 297 1 US-08-534-910B-7 Sequence 7, Appli
37 147.5 8.6 297 1 US-08-534-910B-9 Sequence 9, Appli
38 146.5 8.5 297 1 US-08-534-910B-10 Sequence 10, Appli
39 146.5 8.5 297 3 US-08-886-466-2 Sequence 2, Appli
40 146.5 8.5 297 3 US-09-475-304-2 Sequence 2, Appli
41 146.5 8.5 297 3 US-09-101-126-3 Sequence 3, Appli
42 146.5 8.5 297 3 US-09-367-528A-5 Sequence 5, Appli
43 146 8.5 291 4 US-09-583-110-4604 Sequence 4604, Ap
44 146 8.5 297 1 US-08-534-910B-6 Sequence 6, Appli
45 145.5 8.5 260 4 US-09-710-279-224 Sequence 224, App

ALIGNMENTS

RESULT 1

US-08-761-344-2
; Sequence 2, Application US/08761344
; Patent No. 5912154
; GENERAL INFORMATION:
; APPLICANT: Ferro-No. 5912154ick, Susan
; APPLICANT: Jiang, Yu
; TITLE OF INVENTION: GERANYLGERANYL DIPHOSPHATE SYNTHASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/761,344
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 30,020
REFERENCE/DOCKET NUMBER: 3161-14
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/862-0223
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-761-344-2

Query Match 100.0%; Score 1720; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAKIDELINNDPVWSSQNESLISKPYNHLLKPKGNFRNLIVQINRVNMLPKDQLAIV 60
Db 1 MEAKIDELINNDPVWSSQNESLISKPYNHLLKPKGNFRNLIVQINRVNMLPKDQLAIV 60
QY 61 SQIVELLHNSLLIIDDIEDNAPLRGQTTSHLIFGVFPTINTANYMYFRAMQVLSQLFTK 120

Db 69 NSKLRGFPVAHSYIGIPSVINSANYVFLGLEKV--LTLDP--DAVKLFTQLLELH 123
Qy 140 RQGLDIYWRDFLEIPIPTQEMYLNMVNMKTGGLFRLTLRLMEALSPSSHGHSLVFPIN 199
Db 124 QGGLDIYWRD--NYTPEEBEYKAWVLQKTGGLFGLAVGLNQLF---SDYKEDLKPLN 178
Qy 200 LLGIYIQRDDYLNKDFQMSSEKGFADITEGKLSFPIVHALNFTKTKGOTEQHEILR 259
Db 179 TLGIFQIRDDYANLHSEYSEKSLGDLTEGKFSFTIHAI-----WSRSESTQVQN 232
Qy 260 ILLRTSDKDKLKLQILEFTDTSNLAITKYNFQINOLVNMKNENKYLPLASHSDTAT 319
Db 233 ILKQRTENIDIKKYCVHYLE-DVSGSEYTRNTLKL-----EAK-----AYKOIDAR 278
Qy 320 NHDLELYIIDHSEL 335
Db 279 GGNPELVALVXHL SKN 294

RESULT 3
US-09-038-596-2
; Sequence 2, Application US/09038596
; Patent No. 5928924
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/038,596
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1744
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-038-596-2
Query Match 33.4%; Score 574; DB 2; Length 300;
Best Local Similarity 40.2%; Pred. No. 7.4e-47;
Matches 127; Conservative 55; Mismatches 104; Indels 30; Gaps 8;

Qy 20 ESLISKPNHLLPGKGNFRNLIVQINRVNLPKDLAIVSQIVELHNSLLIDDED 79
Db 9 QRILLEPKYLLQLPGQVTKLSQAFNHLKVPEDKLIITVEMLHNASLLIDDED 68

Db 61 SQIVELHNSLLIDDEDNAPLRGQTSLSHIFGVPSNTANTMYFRAMQLVSLITK 120
Qy 121 EPLVHNLITIFNEELINLHRQGLDIYWRDFLEIPIPTQEMYLNMVNMKTGGLFRLTLRL 180
Db 121 EPLVHNLITIFNEELINLHRQGLDIYWRDFLEIPIPTQEMYLNMVNMKTGGLFRLTLRL 180
Qy 181 MEALSPSSHGHSLVFPINLLGIYIQRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPINLLGIYIQRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Qy 241 ALNFTKTKGOTEQHEILRLILRLSDKDKLKLQILEFTDTSNLAITKYNFQINOLVNMKN 300
Db 241 ALNFTKTKGOTEQHEILRLILRLSDKDKLKLQILEFTDTSNLAITKYNFQINOLVNMKN 300
Qy 301 NDENKYLPLASHSDTATNLDHLELYIIDHSEL 335
Db 301 NDENKYLPLASHSDTATNLDHLELYIIDHSEL 335

RESULT 2
US-08-469-665-2
; Sequence 2, Application US/08469665
; Patent No. 5786193
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-469-665-2
Query Match 33.4%; Score 574; DB 1; Length 300;
Best Local Similarity 40.2%; Pred. No. 7.4e-47;
Matches 127; Conservative 55; Mismatches 104; Indels 30; Gaps 8;

Qy 20 ESLISKPNHLLPGKGNFRNLIVQINRVNLPKDLAIVSQIVELHNSLLIDDED 79
Db 9 QRILLEPKYLLQLPGQVTKLSQAFNHLKVPEDKLIITVEMLHNASLLIDDED 68
Qy 80 NAPLRGQTSLSHIFGVPSNTANTMYFRAMQLVSLITKPEPLYHNLITIFNEELINLH 139

115 -----SOLTKRPLYNLITIFNEELINLHRCGLDLYVRD 150
134 QSSVSSASSSSASSENGGTPTNSQIPFSDTY--LDKVTDEMLSHRGGLFLWRD 191
151 FLPEIIPTEQMYLNMVNKTLGLRLTLRLMEALSPSSHGHSLVPFNLGLIYQIRDD 210
192 SL--TCPSEEFYKVMVGKTLGFLFIAVRLMA---KSECIDFFQVNLVLSIYFQIRDD 246
211 YLNKDFQMSSEKGFPAEDITEKLSPPVHALNFTKQTEQHNEIRILLRTSDKDI 270
247 YWNLQSSEYAHKNFAEDITEKLSFPPTIHSIHANPSS-----RLVINTLQKKSTSEI 300
271 KKLQILEFDNLSLAYTKNFNLVNMKNKND-----NENKYLPLDLASHSDRAT 319
301 LHCHVNYRTETHSFTEQVLTNLISGALRELGRLOGEFAFANSRMDLGDVDSRGTKK 360
320 NLHDELLYIIDLSEL 335
361 NV--KLEALLKKLADI 374

Query Match 14.1%; Score 242.5; DB 4; Length 307;
Best Local Similarity 25.2%; Pred. No. 4.5e-15;
Matches 77; Conservative 67; Mismatches 132; Indels 29; Gaps 8;

4 KIDELINNDPVWSSONESLISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLAIVSQI 63
21 RVDEALNQAQLTLPFSDMPLSQAMRYGALGGKRLRPFLVYAVGEMFNVPVANLDVFAA 80
64 VELLHNSLLIDIE--DNAPLRGQTTSHLIFGVPSTINTANYMYFRAMQVLSOLTKE 121
81 IECIHAYSLHDDLPAWMDNDRGKPTCHIEFGEANILAGDALQTLAEFIIAKNAMPD 140
122 PLYNHLITFNE-----ELINLHRCGLDIYWRDFLPEIIPTEQMYLNMVNKTKGFLRL 176
141 VATADRVMATKATASGLAGMCGQALDDAEDKSIDVALEKHL---HKTGALIRA 196
177 TLRMEALSPSSHGHSLVP-----FINLGIYQIRDDYNLKDQMSSEKGFPAEDITEG 232
197 AVRL--GALS--AGQKGDVLPALDKYAHISGLAQVQDDILDVIGSTGTEETGRQSGDEAG 254
233 KLSFPVHALNFTKQTEQHNEIRILLRTSDKDIKLQILEFDNLSLAYTKNF 292
255 KSTYPALLGLAQAKQKAO--ELYNEALD-----ALAFLEQYEDYSTLKQLANFI 302
293 NQLVN 297
303 VERKN 307

Query Match 13.5%; Score 233; DB 4; Length 341;
Best Local Similarity 21.5%; Pred. No. 4.3e-14;
Matches 65; Conservative 81; Mismatches 135; Indels 22; Gaps 6;

22 LISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLAIVSQIIVLHNSLLIDIEDNA 81
50 LVEKIGDVIISAGGRRLRPLLVLLAGKTILGYKGGDDLCLIAATIEFLHTSTLLHDDVDAS 109
82 FLRRQYTSHLIFGVPSTINTANYMYFRAMQVLSOLTKEPLYNHLITIFNEELINLHRC 141
110 GLRRGRSTANALWGNAPSVLVGDFLYARSFEMVGLSGMP-----VVRILSQATRVIAEG 164
142 QGLDIYWRDFLPEIIPTEQMYLNMVNKTKGFLRLTLRLMEAL--SPSSHGHSLVPFNL 200
165 EVLQL---SKVDDASTTEYMEVIRGKTAMLFEAETHSAALCOAGEHQSEALRRFGDY 221
201 LGIYQIRDDYNLKDQMSSEKGFPAEDITEKLSFPVHALNFTKQTEQHNEIRIL 260
222 LGIAFQVDDLLDYRGDAATLGNVGGDDLAEGKPTLPLIVTM-----RGTBEQAAALVRK 276
261 LLRLTSDKDIK-----LKLQILEFDNLSLAYTKNFNLVNMKNKNDENKYLPLDLASHS 315
277 ALQGGSQDLSEVCAAVEAAGALDYTN---LARDYAARAIACLDLTPDNEYRSALVELS 333
316 DTA 318
334 BFA 336

Query Match 13.5%; Score 232; DB 4; Length 165;
Best Local Similarity 36.8%; Pred. No. 1.9e-14;

; Sequence 21448, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21448
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21448

Query Match 13.5%; Score 233; DB 4; Length 341;
Best Local Similarity 21.5%; Pred. No. 4.3e-14;
Matches 65; Conservative 81; Mismatches 135; Indels 22; Gaps 6;

22 LISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLAIVSQIIVLHNSLLIDIEDNA 81
50 LVEKIGDVIISAGGRRLRPLLVLLAGKTILGYKGGDDLCLIAATIEFLHTSTLLHDDVDAS 109
82 FLRRQYTSHLIFGVPSTINTANYMYFRAMQVLSOLTKEPLYNHLITIFNEELINLHRC 141
110 GLRRGRSTANALWGNAPSVLVGDFLYARSFEMVGLSGMP-----VVRILSQATRVIAEG 164
142 QGLDIYWRDFLPEIIPTEQMYLNMVNKTKGFLRLTLRLMEAL--SPSSHGHSLVPFNL 200
165 EVLQL---SKVDDASTTEYMEVIRGKTAMLFEAETHSAALCOAGEHQSEALRRFGDY 221
201 LGIYQIRDDYNLKDQMSSEKGFPAEDITEKLSFPVHALNFTKQTEQHNEIRIL 260
222 LGIAFQVDDLLDYRGDAATLGNVGGDDLAEGKPTLPLIVTM-----RGTBEQAAALVRK 276
261 LLRLTSDKDIK-----LKLQILEFDNLSLAYTKNFNLVNMKNKNDENKYLPLDLASHS 315
277 ALQGGSQDLSEVCAAVEAAGALDYTN---LARDYAARAIACLDLTPDNEYRSALVELS 333
316 DTA 318
334 BFA 336

Query Match 13.5%; Score 232; DB 4; Length 165;
Best Local Similarity 36.8%; Pred. No. 1.9e-14;

[illegible]

RESULT 10
US-08-898-560-1
; Sequence 1, Application US/08898560
; Patent No. 5935832
; GENERAL INFORMATION:
; APPLICANT: HIROYUKI NAKANE, Chikara OHTO, Shinichi OHNUMA,
APPLICANT: KAZUTAKE HIROOKA, TOKUZO NISHINO
; TITLE OF INVENTION: Farnesyl Diphosphate Synthase

```

1 / GENERAL INFORMATION:
2 / APPLICANT: HIROYUKI NAKANE, Chikara OHTO, Shinichi OHNUMA,
3 / APPLICANT: Kazutake HIROOKA, Tokuzo NISHINO
4 / TITLE OF INVENTION: Parnesyl Diphosphate Synthase
5 / NUMBER OF SEQUENCES: 14
6 / CORRESPONDENCE ADDRESS:
7 / ADDRESSEE: Kenyon & Kenyon
8 / STREET: One Broadway
9 / CITY: New York
10 / STATE: NY
11 / COUNTRY: USA
12 / ZIP: 10004
13 / COMPUTER READABLE FORM:
14 / MEDIUM TYPE: 3+ Floppy disk
15 / COMPUTER: IBM PC compatible
16 / OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
17 / SOFTWARE: WordPerfect 6.1 Windows
18 / CURRENT APPLICATION DATA:
19 / APPLICATION NUMBER: US/08/898,560
20 / FILING DATE: Concurrent Herewith
21 / CLASSIFICATION: 435
22 / PRIOR APPLICATION DATA:
23 / APPLICATION NUMBER: JP 8-213211
24 / FILING DATE: 24-JUL-96
25 / ATTORNEY/AGENT INFORMATION:
26 / NAME: Edward W. Greason
27 / REGISTRATION NUMBER: 18,918
28 / REFERENCE/DOCKET NUMBER: 77670/495
29 / TELECOMMUNICATION INFORMATION:
30 / TELEPHONE: (212) 425-7200
31 / TELEFAX: (212) 425-5288
32 / INFORMATION FOR SEQ ID NO: 1:
33 / SEQUENCE CHARACTERISTICS:
34 / LENGTH: 330 amino acids
35 / TYPE: amino acid
36 / TOPOLOGY: linear
37 / MOLECULE TYPE: protein
38 / ORIGINAL SOURCE:
39 / ORGANISM: Sulfolobus acidocaldarius
40 / STRAIN: ATCC 33909
41 / FEATURE:
42 / NAME/KEY: Asp-rich domain
43 / LOCATION: 82-86
44 / US-08-898-560-1

```

Query Match 13.3%; Score 228.5; DB 1; Length 330;
Best Local Similarity 25.8%; Pred. No. 1.1e-13;
Matches 83; Conservative 65; Mismatches 151; Indels 23

```

Query Match      13.3%; Score 228.5; DB 2; Length 330;
Best Local Similarity 25.8%; Pred. No. 1.1e-13;
Matches 83; Conservative 65; Mismatches 151; Indels 23; Gaps 10;

QY 6 DELINN-DPVWSSQNESLISKPN---HLLKPGKFNRLNLIQVNRVNNLPKQDLAIIV 61
Db 9 NEIVNSVNDIKSYISGDPVKLYEASYHLFTSGGKRLPLIITISSDLFGGQREYAYAG 68
QY 62 QIVELLHNSLLIDIEDNAPLRGGTSHLIFGVPSSTINTANYMYFRAMQIVSOLTXX 121
Db 69 AAIEVLHFTLVHDDIMDQNNRRGLPTVHVYGLPLAILAGDLHLAKAFQLLTQALRGL 128
QY 122 P--LYHNLITIFNEELNLRHGGGLDIYWRDPLPEIPTQEMVYNNVNTGGGLFRLLTLR 179
Db 129 PSETIKAFDIFTRIIISBQAVDMFEDRID--IKEQE-YLDMISRTAALFSASS 185
QY 180 LMEALSPSHHGHSLV-PFINLIGIYQIRDDYLNKDFQMSSEKGFAPDITEGKLSFPI 238
Db 186 IGALTAGANDVRLMSDEGTNLGIAFQIVDDILGLTADKELGKPVFSDIREGKKTILV 245
QY 239 VHALNFTKYGTEQHNEILRIILLRTSDKDKLKIQLIEPDTNSLAYT-----KNFI 292
Db 246 IKTELCK---EDEKKIVLKALGNKSKASKEELMSADIKKYSLDYAYNAEKYKNAI 301
QY 293 NOLVNI--KDNENKYLPLDA 312
Db 302 DSL-NOVSSKSDIPGKALKYLA 322

RESULT 11
US-09-489-039A-13653
; Sequence 13653, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13653
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13653

Query Match      13.2%; Score 227; DB 4; Length 342;
Best Local Similarity 25.4%; Pred. No. 1.6e-13;
Matches 67; Conservative 64; Mismatches 105; Indels 30; Gaps 9;

QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLAIIV 60
Db 36 VNAITLQLNSD-----VQLNQLGYIVSGGKRIKPIAIVLAARAVGYQGSAAHVTI 88
QY 61 SQIVELLHNSLLIDIEDNAPLRGGTSHLIFGVPSSTINTANYMYFRAMQIVSOLTXX 120
Db 89 AALIEFTHTATLHDDYVDSNRRGKATANAFAAGNAASVLVGDFTYTRAFQMMTQLGSL 148
QY 121 EPIYHNLITIFNEELNLRHGGGLDIY-WRDPLPEIPTQEMVYNNVNTGGGLFRLLTLR 179
Db 149 K-----ILEWSEAVNVIAEGFVQLMNVD--PDI--TEENYRVYISKARLFENASQ 199
QY 180 LMEALSP-SHHGHSLVFPFINLIGIYQIRDDYLNKDFQMSSE---KGFAPDITEGKLS 235
Db 200 CAGLLADCTAEERALQDYKRYLGTAFQLIDLLI---DYSSDGBELKNGVGDIDNEGKPT 256
QY 236 FPIVHALNF-----TYTKGQTEQHN 255
Db 257 LPLIHAMHGHGTPDOSAMIRGAIBQGN 282

```

```

RESULT 12
US-09-217-609A-2
; Sequence 2, Application US/09217609A
; Patent No. 6071733
; GENERAL INFORMATION:
; APPLICANT: MURAMATSU, Masayoshi
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OGURA, Kyozi
; APPLICANT: Koyama, Tanetoshi
; APPLICANT: SHIMIZU, Naoto
; APPLICANT: CHO, Yenwin
; TITLE OF INVENTION: Prenyl Diphosphate Synthetase Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, NW - Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/217,609A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/873,235
; FILING DATE: 11-Jun-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: TOFFENETTI, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 10235/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-217-609A-2

Query Match      13.1%; Score 225.5; DB 3; Length 325;
Best Local Similarity 24.3%; Pred. No. 2.1e-13;
Matches 81; Conservative 72; Mismatches 144; Indels 37; Gaps 11;

QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLAIIV 60
Db 17 VEKRYECIQSD-----SETINKAAHLLSSGGKRVPMFVLLSGFLNDTQKDLIRT 69
QY 61 SQIVELLHNSLLIDIEDNAPLRGGTSHLIFGVPSSTINTANYMYFRAMQIVSOLTXX 120
Db 70 AVSLELVHVASLVHDDYIDNSDMRGNTSVHIAFDKDTALRTGHFLARALQNIATNNS 129
QY 121 EPIYHNLITIFNEELNLRHGGGLDIYWRDPLPEIPTQEMVYNNVNTGGGLFRLLTLR 180
Db 130 K--FHQ---IFSXTILEVCFGEFDQADRFNYPVSFTA---YLERINRKTALIEASCHL 181
QY 181 MEALS-----PSSHGHSLVFPFINLIGIYQIRDDYLNKDFQMSSEKGFAPDITEGKLS 235
Db 182 -GALSQLEBQSYH---IKQGHCIQMSYQIIDDIDYTSDEATLGKPVGSIRNGHIT 237
QY 236 FPIVHALNFTKYGTEQHNEILRIILLRTSDKDKLKI-QILEPDTNSLAYTQNFQ 294
Db 238 YPLMAAI--ANLKEQDDDKLEAVVGHILTSTSDDEVYQIVSVQKYGIEPA-----E 287

```

QY 295 LVNMIKNDNENKYLDEL--ASHSDTATNLHDELL 326
Db 288 LLSRKYGDKAKYHLSQLQDSNLIKDYLEIHEKML 321

RESULT 13

US-08-873-235B-2
; Sequence 2, Application US/08873235B
; Patent No. 6174715
; GENERAL INFORMATION:
; APPLICANT: MURAMATSU, Masayoshi
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OGURA, Kyozi
; APPLICANT: KOYAMA, Tanetoshi
; APPLICANT: SHIMIZU, Naoto
; APPLICANT: CHO, Yenwin
; TITLE OF INVENTION: Prenyl Diphosphate Synthetase Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, NW - Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: Wordperfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,235B
; FILING DATE: 11-Jun-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 154441/1996
; FILING DATE: 14-Jun-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: TOFFENETTI, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 10235/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-873-235B-2

Query Match 13.1%; Score 225.5; DB 3; Length 325;
Best Local Similarity 24.3%; Pred. No. 2.1e-13;
Matches 81; Conservative 72; Mismatches 144; Indels 37; Gaps 11;
QY 1 MEAKIDELINNDPWSQNESLISKYPN--HILLKPKNFRNLIVQINRMVNLKQDLAIV 60
Db 17 VEKRLVEICQSD-----SETINKAAHILSSGGRKVRPMFVLLSGFLNDTQKDLIRT 69
QY 61 SQIVELLHNSLLDIEDNAPLRGQTTSHLIFGVPTSTINTANYMYFRAMQLVSLTTK 120
Db 70 AVSLELVHMASLVHDDYIDNSDMERGNTSVHIAFDKDTARTGHFLARALQNIATINNS 129
QY 121 EPLVHNLIITFNEELINLHRGQGLDIYWRDFLEIPIPTQEMVNLNMVNTKGGFLRLTLR 180
Db 130 K--FHQ---IFSKTILEVCFGEFDQMDRPNYFVSFTA---YLRRINRKTALLIERASCHL 181
QY 181 MEALS-----PSSHHGSHLV--PFNLIGIYQIRDDYINLKDQFMSSEKGFADITEGKLSFPI 235
Db 182 -GALSSQLDQSYTH---IKQFGHCIGMSYQIIDDILDYTSDEATLGKPVGSDIRNGHIT 237
QY 236 FPIVHALNFTTKGQTEQHNEILRIILLRTSDKDIKLLI-QILEPDTNSLAYTKMFINQ 294

Db 238 YPLMAAI--ANLKEQDDDDKLEAVVKHLTSTSDDEVYQIVSQVKYGLEPA-----E 287
QY 295 LVNMIKNDNENKYLDEL--ASHSDTATNLHDELL 326
Db 288 LLSRKYGDKAKYHLSQLQDSNLIKDYLEIHEKML 321

RESULT 14

US-09-101-126-1
; Sequence 1, Application US/09101126
; Patent No. 6316216
; GENERAL INFORMATION:
; APPLICANT: OHTO, CHIKARA
; APPLICANT: NAKANE, HIROYUKI
; APPLICANT: NISHINO, TOKUZO
; APPLICANT: OHNUMA, SHINTICHI
; APPLICANT: HIROOKA, KAZUTAKE
; TITLE OF INVENTION: MUTATED PRENYL DIPHOSPHATE SYNTHASES
; FILE REFERENCE: 77670/566
; CURRENT APPLICATION NUMBER: US/09/101,126
; CURRENT FILING DATE: 1999-04-27
; EARLIER APPLICATION NUMBER: PCT/JF97/03921
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: JP 8-307506
; EARLIER FILING DATE: 1996-11-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Sulfolobus acidocaldarius
; FEATURE:
; OTHER INFORMATION: 82-86 is an Asp-rich domain
; US-09-101-126-1

Query Match 13.1%; Score 225.5; DB 3; Length 330;
Best Local Similarity 25.8%; Pred. No. 2.1e-13;
Matches 83; Conservative 64; Mismatches 152; Indels 23; Gaps 10;
QY 6 DELINN-DPVWSSQNESLISKYPN---HILLKPKNFRNLIVQINRMVNLKQDLAIVS 61
Db 9 NEIVNSVNDIISKVSDVPKLYEASVHLFTSGKRLRPILITSSDLFGQGRAYYAG 68
QY 62 QIVELLHNSLLDIEDNAPLRGQTTSHLIFGVPTSTINTANYMYFRAMQLVSLTTK 121
Db 69 AAIEVLHTFTLVHDDIMDQDNIRRGPLTVHVKYGLPIAILAGLLHAKAFQLLTQALRGL 128
QY 122 P--LYHNLIITFNEELINLHRGQGLDIYWRDFLEIPIPTQEMVNLNMVNTKGGFLRLTLR 179
Db 129 PSETIIKAFDIFTRSIISSEGOAVDMFEFDRID--IKEQE-YLDMISRKTALFSASS 185
QY 180 LMEALSPSSHGHSHLV--PFNLIGIYQIRDDYINLKDQFMSSEKGFADITEGKLSFPI 238
Db 186 IGALIAAGANDVRLMSDFCTNIGIAFQIVDDIILGTADEKELGKPVFSDIRGKKTILV 245
QY 239 VHALNFTTKGQTEQHNEILRIILLRTSDKDIKLLIQLIEPDTNSLAYT-----KNFTI 292
Db 246 IKTLELCK----EDEKKIVLKALGNKSKASKEELMSADIKKYSLDYAVNLAEEKYKNAI 301
QY 293 NQLVNMII--KNDMENKYLPLDA 312
Db 302 DSL-NOVSSKSDIPGKALKYLA 322

RESULT 15

US-09-543-681A-5778
; Sequence 5778, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5778
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5778

Query Match      12.9%; Score 221.5; DB 4; Length 340;
Best Local Similarity 23.1%; Pred. No. 5.4e-13;
Matches 68; Conservative 71; Mismatches 114; Indels 41; Gaps 8;

Qy      11 NDPVSSQNE--SLISKPNHILLKPGKNFRNLIVQINRVNMLPKDQIAIVSQIVELLH 68
Db      11 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      35 NDAILYQLNSDVSLLINQLGYIISGGKRIIRPMIAVLTGRALDYQGDKHISVAALIEFIH 94
Db      35 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      69 NSSLIIDIEDNAPLRGQTTHLIFGVPSITANTANYMYFRAMQLVSQLTKKEPIYHNLJ 128
Db      69 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      95 TATLHDDVDVDESDMRGKQTANAVFGNAASVLVGDFTYTRSFQMTDLDNSRVL--KLM 152
Db      95 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      129 T-----IFNEELINLHRGQGLDIYWRDPLPEIIPQEMVLYNNMKNKGTGLFRLRLMEA 183
Db      129 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      153 SSATNVIAGEVQLMNCNDPDI-----TEDDYMQVIYSKTA-----RLFEA 194
Db      153 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      184 LSPSS-----HGHSLVPPFINLLGIYQIRDDYLNKDPQMSSEKGFADITEGKLS 235
Db      184 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      195 ASHASAILCGATPEQEKAFQDYGRVLTAFQLIDLDLDYDADNTQLGKNTGDDLDGKPT 254
Db      195 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      236 PPIVHALNFTKQTEQHNEILRIILLRTSDKDIKLIQILEFDTSLAYTK 289
Db      236 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Qy      255 LPLHAMHH-----GNETESQLIROAIEKGNRHLDDVLTWK--QCGSLEYTR 302
Db      255 :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
```

Search completed: October 17, 2004, 08:56:04
Job time : 29.4406 secs

| | Query Match | 100.0.0%; | Score 1569; | DB 16; | length 1569; | |
|----|-----------------------|---|----------------------------|------------|--------------|--------|
| | Best Local Similarity | 100.0.0%; | Pred. No. 0; | | | |
| | Matches 1569; | Conservative | 0; | Mismatches | 0; | Indels |
| | | | | | | Gaps |
| QY | 1 | AATATTACATATAGATATAGGACAAGCCCGCATTTTCATAC | CTGAAAGGTAAACTTC | TATT | 60 | |
| Db | 1 | AATATTACATATAGATATAGGACAAGCCCGCATTTTCATAC | CTGAAAGGTAAACTTC | TATT | 60 | |
| QY | 61 | ATTATAGTGGTATCCAAAGCTTCA | CCGCTTCCAGCATAGCAGAAAT | TACG | TGTTTTTGCAT | 120 |
| Db | 61 | ATTATAGTGGTATCCAAAGCTTCA | CCGCTTCCAGCATAGCAGAAAT | TACG | TGTTTTTGCAT | 120 |
| QY | 121 | ATGTTATGCTGAATCATTTGATGCT | TAACCAATTTTCTTTTGCTTGCCTTC | TCTTT | 180 | |
| Db | 121 | ATGTTATGCTGAATCATTTGATGCT | TAACCAATTTTCTTTTGCTTGCCTTC | TCTTT | 180 | |

QY 181 GACGTTTGTGAGCAAAAAAGTCAAGACAGATGTGCTTACAAAACCAATGTAAGGC 240
Db 181 GACGTTTGTGAGCAAAAAAGTCAAGACAGATGTGCTTACAAAACCAATGTAAGGC 240
QY 241 TCAATTTTCAAGAAGCTACTAATAGAAAGAGAACAAAGAGTTTACGAGTCTGGAATCA 300
Db 241 TCAATTTTCAAGAAGCTACTAATAGAAAGAGAACAAAGAGTTTACGAGTCTGGAATCA 300
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTGTGTTGGTCCAGCCAAAATGAA 360
Db 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTGTGTTGGTCCAGCCAAAATGAA 360
QY 361 AGCTTGATTTCAAAACCTTATATACATCTCTTTGAAACCTTGGTCCAGCAACTTTAGACTA 420
Db 361 AGCTTGATTTCAAAACCTTATATACATCTCTTTGAAACCTTGGTCCAGCAACTTTAGACTA 420
QY 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT 480
Db 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT 480
QY 481 TCGCAAAATTTGAGCTCTTGCAATAATCCAGCCTTTTAAATCGACGATATAGAAGATAAT 540
Db 481 TCGCAAAATTTGAGCTCTTGCAATAATCCAGCCTTTTAAATCGACGATATAGAAGATAAT 540
QY 541 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
Db 541 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
QY 601 AACACCGCAAAATATATATGATTTTACAGAGCCATCAACTTGTATCGCAGCTTAACCAAAA 660
Db 601 AACACCGCAAAATATATATGATTTTACAGAGCCATCAACTTGTATCGCAGCTTAACCAAAA 660
QY 661 GAGCCTTTGATCATAAATTTGATTAAGATTTTCAAGAGAAATGATCAATCTACATAGG 720
Db 661 GAGCCTTTGATCATAAATTTGATTAAGATTTTCAAGAGAAATGATCAATCTACATAGG 720
QY 721 GGAAGAGCTTGATATATATGATTTTACAGAGCCATCAACTTGTATCGCAGCTTAACCAAAA 780
Db 721 GGAAGAGCTTGATATATATGATTTTACAGAGCCATCAACTTGTATCGCAGCTTAACCAAAA 780
QY 781 ATGATTTTGAATATGTTGATTAAGATTTTCAAGAGAAATGATCAATCTACATAGG 840
Db 781 ATGATTTTGAATATGTTGATTAAGATTTTCAAGAGAAATGATCAATCTACATAGG 840
QY 841 ATGGAAGCGCTGTCTCTCTCAACACCGGCAATCGTGTGTTTCTTTCATTAATCTT 900
Db 841 ATGGAAGCGCTGTCTCTCTCAACACCGGCAATCGTGTGTTTCTTTCATTAATCTT 900
QY 901 CTGGGTATTTATATGATTTAGATGATTTTCAAGAGAAATGATCAATCTACATAGG 960
Db 901 CTGGGTATTTATATGATTTAGATGATTTTCAAGAGAAATGATCAATCTACATAGG 960
QY 961 AGCGAAAAGGCTTGTGCTGAGGACATTTACAGAGGGGAGTTATCTTTTCCCATCGTCCAC 1020
Db 961 AGCGAAAAGGCTTGTGCTGAGGACATTTACAGAGGGGAGTTATCTTTTCCCATCGTCCAC 1020
QY 1021 GCCTTTAACTTCACTAAAACGAAAGGTCAAACTGAGCAACAAATGAAATTTCTAAGAAAT 1080
Db 1021 GCCTTTAACTTCACTAAAACGAAAGGTCAAACTGAGCAACAAATGAAATTTCTAAGAAAT 1080
QY 1081 CTCTGTTGAGGACAAAGTATAGATATATAAATCAAGCTGATTTCAATCTGGAATTC 1140
Db 1081 CTCTGTTGAGGACAAAGTATAGATATATAAATCAAGCTGATTTCAATCTGGAATTC 1140
QY 1141 GACACCAATTCATTTGGCCCTACACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200
Db 1141 GACACCAATTCATTTGGCCCTACACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200
QY 1201 AATGATATAAATAAGTATTTTACCTGATTTGGCTTGGCATTTCCGACACCGCCACCAAT 1260
Db 1201 AATGATATAAATAAGTATTTTACCTGATTTGGCTTGGCATTTCCGACACCGCCACCAAT 1260

QY 1261 TTACATGACGAATTTGTTATATATATATAGACCACTTATCCGAATTTGTAATAAATGATC 1320
Db 1261 TTACATGACGAATTTGTTATATATATATAGACCACTTATCCGAATTTGTAATAAATGATC 1320
QY 1321 AATCAAAATAGTGGAGGAGATAGTCAGAAATAAGCCCTTCTCTCTCTCTCTCTCTCTCT 1380
Db 1321 AATCAAAATAGTGGAGGAGATAGTCAGAAATAAGCCCTTCTCTCTCTCTCTCTCTCTCT 1380
QY 1381 TATACATACGATTTTCATATATACGTTTCAATTCATCATCTTTTGTATATATCTCAAAAAGA 1440
Db 1381 TATACATACGATTTTCATATATACGTTTCAATTCATCATCTTTTGTATATATCTCAAAAAGA 1440
QY 1441 TCTCTTAGTTTCCAAATAGTCAAAATCTTCAAAATTTATATAGCTTTTATATTTTCCACGAT 1500
Db 1441 TCTCTTAGTTTCCAAATAGTCAAAATCTTCAAAATTTATATAGCTTTTATATTTTCCACGAT 1500
QY 1501 TTCTGAAACTCTTTTATATACGACCGTTAATGCTAGCGGTTACTGTCAAAATCGCGGTA 1560
Db 1501 TTCTGAAACTCTTTTATATACGACCGTTAATGCTAGCGGTTACTGTCAAAATCGCGGTA 1560
QY 1561 AATTCGCGA 1569
Db 1561 AATTCGCGA 1569
RESULT 2
US-10-369-493-46181
; Sequence 46181, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46181
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; US-10-369-493-46181
Query Match 64.2%; Score 1008; DB 15; Length 1008;
Best Local Similarity 100.0%; Pred. No. 5.6e-230;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTTTGGTCCAGCCAAAATGAA 360
Db 1 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTTTGGTCCAGCCAAAATGAA 60
QY 361 AGCTTGATTTCAAAACCTTATATACATCTTTTGAACCTGGCAAGAACTTTAGACTA 420
Db 61 AGCTTGATTTCAAAACCTTATATACATCTTTTGAACCTGGCAAGAACTTTAGACTA 120
QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCAAGCTGGCCATAGTT 480
Db 121 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCAAGCTGGCCATAGTT 180
QY 481 TCGCAAAATTTGAGCTCTTGCATATTTCCAGCCTTTTAAATCGACGATATAGAAGATAAT 540
Db 181 TCGCAAAATTTGAGCTCTTGCATATTTCCAGCCTTTTAAATCGACGATATAGAAGATAAT 240
QY 541 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
Db 241 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 300

```

QY 601 AACACGGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 660
Db 301 AACACGGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 360
QY 661 GAGCCTTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 720
Db 361 GAGCCTTTGATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 420
QY 721 GGACAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 780
Db 421 GGACAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 480
QY 781 ATGTAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 840
Db 481 ATGTAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 540
QY 841 ATGGAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 900
Db 541 ATGGAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 600
QY 901 CTGGGTATTTATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 960
Db 601 CTGGGTATTTATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 660
QY 961 AGCGAAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 1020
Db 661 AGCGAAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 720
QY 1021 GCCCTTAACTTCACTAAACCAAGGTCAAATGAGCACTAACCAAGGTCAAATGAGCACTAACCAAGGTC 1080
Db 721 GCCCTTAACTTCACTAAACCAAGGTCAAATGAGCACTAACCAAGGTCAAATGAGCACTAACCAAGGTC 780
QY 1081 CTCCTGTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTGATTAAGCTG 1140
Db 781 CTCCTGTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTGATTAAGCTG 840
QY 1141 GACACCAATTCATTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTG 1200
Db 841 GACACCAATTCATTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTG 900
QY 1201 AATGATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGT 1260
Db 901 AATGATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGT 960
QY 1261 TTACATGACGAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 1308
Db 961 TTACATGACGAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 1008

```

RESULT 3

```

US-10-450-941-5
; Sequence 5, Application US/10450941
; Publication No. US20040063182A1
; GENERAL INFORMATION:
; APPLICANT: TOKYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: A METHOD OF PRODUCING PRENYLALCOHOL
; FILE REFERENCE: PH-1444PCT
; CURRENT APPLICATION NUMBER: US/10/450,941
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: JP2000-401701
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2000-403067
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2001-282978
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: (1) .. (1005)
US-10-450-941-5
Query Match 64.2%; Score 1008; DB 16; Length 1008;
Best Local Similarity 100.0%; Pred. No. 5.6e-230;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 301 ATGGAGGCGCAAGATAGATGAGCTGATCAATATATGATCTGTTGTTGCTGAGCCAAATGAA 360
Db 1 ATGGAGGCGCAAGATAGATGAGCTGATCAATATATGATCTGTTGTTGCTGAGCCAAATGAA 60
QY 361 AGCTTGATTTCAAAACCTTTAATACATCTCTTTTGAACCTGCGCAAGCACTTTAGACTA 420
Db 61 AGCTTGATTTCAAAACCTTTAATACATCTCTTTTGAACCTGCGCAAGCACTTTAGACTA 120
QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGAGACAGCTGCCATAGTT 480
Db 121 AATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGAGACAGCTGCCATAGTT 180
QY 481 TCGCAAAATTTGAGCTCTTTCGATTAATTTCCAGCTTTTAAATCGAGATATAGAGATAAT 540
Db 181 TCGCAAAATTTGAGCTCTTTCGATTAATTTCCAGCTTTTAAATCGAGATATAGAGATAAT 240
QY 541 GCTCCCTTGAAGGGGACAGACCACTTCTCACTTAATCTTCCGGTGTACCTCCACTATA 600
Db 241 GCTCCCTTGAAGGGGACAGACCACTTCTCACTTAATCTTCCGGTGTACCTCCACTATA 300
QY 601 AACACCGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 660
Db 301 AACACCGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 360
QY 661 GAGCCTTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 720
Db 361 GAGCCTTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 420
QY 721 GGACAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 780
Db 421 GGACAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 480
QY 781 ATGTAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 840
Db 481 ATGTAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 540
QY 841 ATGGAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 900
Db 541 ATGGAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 600
QY 901 CTGGGTATTTATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 960
Db 601 CTGGGTATTTATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 660
QY 961 AGCGAAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 1020
Db 661 AGCGAAAAGGCTTGGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 720
QY 1021 GCCCTTAACTTCACTAAACCAAGGTCAAATGAGCACTAACCAAGGTCAAATGAGCACTAACCAAGGTC 1080
Db 721 GCCCTTAACTTCACTAAACCAAGGTCAAATGAGCACTAACCAAGGTCAAATGAGCACTAACCAAGGTC 780
QY 1081 CTCCTGTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTGATTAAGCTG 1140
Db 781 CTCCTGTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTGATTAAGCTG 840
QY 1141 GACACCAATTCATTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTG 1200
Db 841 GACACCAATTCATTTGAGCAAGTGAATAAGATATAAATCTAAAGCTGATTAAGCTGATTAAGCTG 900
QY 1201 AATGATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGT 1260
Db 901 AATGATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGTGAATAAGT 960
QY 1261 TTACATGACGAATTTGATATATATGTAATTTTCAGAGCCATGCAACTTGTATGCTGAGCACTAACCAAAAA 1308

```

Db 961 TTACATGACGAATCTTTATATATATATAGACCACCTTATCCGAATTGTGA 1008

RESULT 4
US-10-189-268-4
; Sequence 4, Application US/10189268
; Publication No. US20040005570A1
; GENERAL INFORMATION:
; APPLICANT: Nicholas M. Dean
; APPLICANT: C. Frank Bennett
; APPLICANT: Kenneth W. Dobbie
; TITLE OF INVENTION: ANTISENSE MODULATION OF GERANYLGERANYL DIPHOSPHATE SYNTHASE 1 EXH
; FILE REFERENCE: PTS-0021
; CURRENT APPLICATION NUMBER: US/10/189,268
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 131
; SEQ ID NO 4
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (170)...(1072)
US-10-189-268-4

Query Match 7.3%; Score 114; DB 16; Length 1395;
Best Local Similarity 52.1%; Pred. No. 1.2e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAGCTTGGATTTCAAAACCTTATAATACATCCTTTTGAACCTGGCAAGACTTTAGAC 418
Db 195 AAGAAATCTTCTAGAACCTTATAATATCTTACTTACAGTTACAGTAAACAAGTGAGAA 254

QY 419 TAAATTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGTGGCCATAG 478
Db 255 CCAAACTTTACAGGCAATTAATCATTTGGCTGAAAGTTCCAGAGGACAAAGCTACAGATTA 314

QY 479 TTTCGCAAAATTTGAGCTCTTCATTAATCCAGCTTTTAAATCGAGGATATAGAAGATA 538
Db 315 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATTTGAAGACA 374

QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTA 598
Db 375 ACTCAAACTCCGACCTGGCTTTCCAGTGCCCCACAGCATCTATGAATCCCATCTGTCA 434

QY 599 TAAACACCGCAATTAATATATGATTTTCAGAGCCATGCAACTTTGTATCGAGCTAACACAA 658
Db 435 TCAATCTGCCAATTACGTGTATTTCTTGGCTTGGAGAAAGTCT----- 479

QY 659 AAGAGCTTTGTATCATATTTGATTAAGATTTTCAAGAGAAATGATCAATCTACATA 718
Db 480 TAACTCTGATCACCCAGATGCGAGTGAAGCTTTTACCCGCCAGCTTTTGGAACTCAATC 539

QY 719 GGGGCAAGGCTTTGGATATATATCTGAGAGACTTTCTGCCGTGAATCATACCTACTCAGG 778
Db 540 AGGGAACAGCCCTAGATATTTACTGAGGGA-----TAATTAACACTTTGCCACTGAAG 593

QY 779 AGATGATTTGAATATGTTTATGAATAAACAAGCGGCCCTTTTCAGATTAACGTTGAGAC 838
Db 594 AAGAATATAAGCTATGGTCTGCAGAAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC 653

QY 839 TCATGGAAGCGCTCTCTCTCTCTCCCTCACACAGCGCCATTCGTTGGTTTCCTTTTCATAAATC 898
Db 654 TCATGCAAGTTGTTCTCTGATTAACAGAGATTTAAACCGCTACTT-----ATA 704

QY 899 TTCTGGGTATTTATTTACAGATTAGAGATGATTACTTTGAATTTGAAGATTTCCAAATGT 958
Db 705 CACTTGGGCTCTTTTCCAAATTAGGGATGATTATGCTAATCTACACTCCAAAGAAATATA 764

QY 959 CCAGCGAAAAGGCTTTGCTGAGACATTTACAGAGGGGAGATTCTTTTCCCATCTGCC 1018
Db 765 GTGAAACAAAAGTTTTTGTGAAGATCTGACAGAGGGAAGTTCTCATTTCTCTACTATTC 824

QY 1019 ACGCCCTT 1026
Db 825 ATGCTATT 832

RESULT 5
US-09-814-353-21762
; Sequence 21762, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21762
; LENGTH: 1692
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..1692
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21762

Query Match 7.3%; Score 114; DB 10; Length 1692;
Best Local Similarity 52.1%; Pred. No. 1.3e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAGCTTGAATTTCAAAACCTTATAATCAATCCTTTTGAACCTGGCAAGACTTTAGAC 418
Db 280 AAGAATTTCTTAGAACCTTATAATCTTACTTCCAGTTACAGGTAAACAAGTGAGAA 339

QY 419 TAAATTTAATAGTTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCCAGCTGCCATAG 478
Db 340 CCAAACTTTACAGGCAATTTAATCATTTGGCTGAAGTTCCAGAGGACAGCTACAGATTA 399

QY 479 TTTCGCAAAATTTGTGAGCTTTTCATTAATTTCCAGCTTTTAAATCGACGATATAGAAGATA 538
Db 400 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATTTGAAGACA 459

QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTA 598
Db 460 ACTCAAACTCCGACCTGGCTTTCCAGTGGCCCAACAGCATCTATGGAATCCCATCTGTCA 519

QY 599 TAAACACCGCAATTTATATGTTTTCAGAGCCATGCAACTTTGTATCGAGCTAACCAAA 658
Db 520 TCAATTTCCCAATTAAGTATTTCTTGGCTTGGAGAAAGTCT----- 564

QY 659 AAGAGCTTTGTATCATTAATTTGATTAAGATTTTCAACAGAGAAATGATCAATCTACATA 718
Db 565 TAAACCTTGTATCACCAGATGCAAGTCTTTTACCCGCCAGCTTTTGGAACTCCATC 624

QY 719 GGGGACAGGCTTTGGATATATATCTGAGAGACTTTTCTGCCTGAAATCATACCTACTCAGG 778

Db 625 AGGACACAGCCCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 678
 QY 779 AGATGTAATTTGAATATGGTTATGAATAAAACAGCGCGCTTTTCAGATTAAACGTTGAGAC 838
 Db 679 AAGATATAAGCTATGGTCTGCAGAAAACAGGTGACTGTTGGATTAGCAGTAGGTC 738
 QY 839 TCATGAAGCGCTGTCTCTCTCTCTCACACACCGGCCAATTCGTTGGTTCCTTTTCATAAATC 898
 Db 739 TCATGCAAGTTGTTCTCTGATTAACAAGAAGATTTAAACCGCTACTT-----AATA 789
 QY 899 TTCTCGGTATTTATCAGATTAGAGATGATTAATCTTGAATTTGAAGATTTCCAAATGT 958
 Db 790 CACTTGGGCTCTTTTCCAAATTAGGATGATTAATGCTAATCTACACTCCAAAGATATA 849
 QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTAACAGAGGGGAAGTTATCTTTTCCCATCGTCC 1018
 Db 850 GTGAAAACAAAAGTTTCTGAGATCTGACAGAGGGAAGTTCTCAATTTCTACTATTC 909
 QY 1019 AGGCCCTT 1026
 Db 910 ATGCTATT 917

RESULT 6

US-10-041-018-3
 ; Sequence 3, Application US/10041018
 ; Publication No. US2004007232A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsuda, Seiichi P.T.
 ; APPLICANT: Hart, Elizabeth A.
 ; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
 ; FILE REFERENCE: P02080US/10025547
 ; CURRENT APPLICATION NUMBER: US/10/041,018
 ; CURRENT FILING DATE: 2002-01-07
 ; PRIOR APPLICATION NUMBER: US 60/259880
 ; PRIOR FILING DATE: 2001-01-05
 ; NUMBER OF SEQ ID NOS: 413
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 2534
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-041-018-3

Query Match 7.3%; Score 114; DB 16; Length 2534;
 Best Local Similarity 52.1%; Pred. No. 1,6e-16;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
 QY 359 AAAGCTTGATTTCAAACCTTATATCAATCATCTTTTGAACCTGGCAAGAACTTTAGAC 418
 Db 258 AAAGAATTTCTTAGAACCTTATAAATCTTACTTACAGTTACCAAGTAAACAAGTGAGAA 317
 QY 419 TAAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAG 478
 Db 318 CCAACTTTACAGGCAATTTAATCATGCTGGAAGTTCCAGAGACAAGCTACAGATTA 377
 QY 479 TTTCGCAATTTGTTGAGCTCTTGCAATATCCAGCTTTTAAATCGACATATAGAATA 538
 Db 378 TTATTGAAGTGACAGAAATGTTGATATGCGAGTTACTCATCGATGATATTGAAGACA 437
 QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 598
 Db 438 ACTCAAACTCCGACGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGCA 497
 QY 599 TAAACCGCAAAATATATGATTTTCAGAGCCATGCACTTGTATCGAGCTAACCAAA 658
 Db 498 TCAATTCGCAATACGTTGATTTCTTGGCTTTGGAGAAAGTCT-----542
 QY 659 AAGAGCCCTTTGATCATAAATTTGATTTAGATTTTCAACGAGAAATTCATCAATCTACATA 718
 Db 543 TAACCCTTGATACCCAGATGCGAGTGAAGCTTTTACCGGCACTTTTGGAACTCCATC 602
 QY 719 GGGGACAAGGCTTGGATATATATCTGGAGAGACTTTTCTGCTGAAATCATCTACTCAGG 778

Db 603 AGGACACAGCCCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 656
 QY 779 AGATGTAATTTGAATATGGTTATGAATAAAACAGCGCGCTTTTCAGATTAAACGTTGAGAC 838
 Db 657 AAGATATAAGCTATGGTCTGCAGAAAACAGGTGACTGTTGGATTAGCAGTAGGTC 716
 QY 839 TCATGAAGCGCTGTCTCTCTCTCTCACACACCGGCCAATTCGTTGGTTCCTTTTCATAAATC 898
 Db 717 TCATGCAAGTTGTTCTCTGATTAACAAGAAGATTTAAACCGCTACTT-----AATA 767
 QY 899 TTCTCGGTATTTATCAGATTAGAGATGATTAATCTTGAATTTGAAGATTTCCAAATGT 958
 Db 768 CACTTGGGCTCTTTTCCAAATTAGGATGATTAATGCTAATCTACACTCCAAAGATATA 827
 QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTAACAGAGGGGAAGTTATCTTTTCCCATCGTCC 1018
 Db 828 GTGAAAACAAAAGTTTCTGAGATCTGACAGAGGGAAGTTCTCAATTTCTACTATTC 887
 QY 1019 AGGCCCTT 1026
 Db 888 ATGCTATT 895

RESULT 7

US-09-925-302-171
 ; Sequence 171, Application US/09925302
 ; Patent No. US20020044941A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PAL04
 ; CURRENT APPLICATION NUMBER: US/09/925,302
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05918
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 896
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 171
 ; LENGTH: 2529
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-302-171

Query Match 7.2%; Score 113.6; DB 9; Length 2529;
 Best Local Similarity 51.9%; Pred. No. 2e-16;
 Matches 347; Conservative 1; Mismatches 290; Indels 30; Gaps 3;
 QY 359 AAAGCTTGATTTCAAACCTTATATCAATCATCTTTTGAACCTGGCAAGAACTTTAGAC 418
 Db 246 AAAGAATTTCTTAGAACCTTATAAATCTTACTTACAGTTACCAAGTAAACAAGTGAGAA 305
 QY 419 TAAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAG 478
 Db 306 CCAACTTTACAGGCAATTTAATCATGCTGGAAGTTCCAGAGACAAGCTACAGATTA 365
 QY 479 TTTCGCAATTTGTTGAGCTCTTGCAATATCCAGCTTTTAAATCGACATATAGAATA 538
 Db 366 TTATTGAAGTGACAGAAATGTTGATATGCGAGTTACTCATCGATGATATTGAAGACA 425
 QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 598
 Db 426 ACTCAAACTCCGACGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGCA 485
 QY 599 TAAACCGCAAAATATATGATTTTCAGAGCCATGCACTTGTATCGAGCTAACCAAA 658
 Db 486 TCAATTCGCAATACGTTGATTTCTTGGCTTTGGAGAAAGTCT-----530
 QY 659 AAGAGCCCTTTGATCATAAATTTGATTTAGATTTTCAACGAGAAATTCATCAATCTACATA 718
 Db 531 TAACCCTTGATACCCAGATGCGAGTGAAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 590

Db 393 CTCTCTTCTCCGCGGGAATTCGGTCGCACATAGCATCTATGGCACCGCGCAGAGAT 452
QY 600 AAACACCGCAAAATATATATGTTTTCAGAGCGATGCAACTTGATCGCAGCTAACCAAA 659
Db 453 CAAATTCGGCAAACTAGCTTTACTTCTCGCCCTCCAGGAGGTGCAAAAATGAAGATCC 512
QY 660 AGAGCCTTTGTATCAFAAATTTGATTACGATTTTCAACGAAGAATTCATCAATCTACATAG 719
Db 513 GCAGCTATCGAC-----ATATAGCTCCAGGAGCTGCTGAATTTACACAG 557
QY 720 GGGACAGCCTTGGATATATATCTGGAGAGACTTTTCGCCGTGAATCATACCTACTCAGGA 779
Db 558 AGGGCAGGAGCATGGATCTGTTTGGCGAGACACGCTCACTTGT-----CCAAGCGAAGA 611
QY 780 GATGATTTGAATATGTTTATGAATAAACAACAGCGCGCTTTTCAGATTAACGTTGAGACT 839
Db 612 TGAATACTTGGAGATGGTGGGCAACAGACTGGAGGTTTGTTCGGCTAGCTGTGAATTT 671
QY 840 CATGGAAGCGCTGTCTCTCTTCCTCAACCAACGCGCCATTCGTTGGTTCTTCATAAATCT 899
Db 672 GATGCAAG-----CTGAAGCAGCACTGGAAGGACTGTGGCCCTTGTGAATGT 722
QY 900 TCTGGGTATTTATCAGATTTAGATGATTAATTTGAAATTTGAAAGATTTCCAAATGTC 959
Db 723 TTTGGGACTGGTCTTTCAGATATGCGACGACTATCTCAATTTATCGACACGACGTATAC 782
QY 960 CAGCCAAAAGGCTTTGCTGAGGACATTACAGAGGGAGTTATCTTTTCCCATCGTCA 1019
Db 783 CCAGAACAAAGGGCTCTGTGAAGACCTCAGAGGGCAAAATTTTCAATTCCTCATCA 842
QY 1020 GCCTCTT 1026
Db 843 CAGCAAT 849

RESULT 10

US-10-369-493-27908
; Sequence 27908, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27908
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-27908

Query Match 5.8%; Score 90.6; DB 15; Length 1284;
Best Local Similarity 49.6%; Pred. No. 4.4e-11;
Matches 339; Conservative 0; Mismatches 314; Indels 30; Gaps 3;

QY 343 TGGTCCAGCCAAATGAAGCTTGATTTCAAAACCTTATATCATCATCTTTTGAACCT 402
Db 361 TGGTCCGAAGAGAGGAGAGGTTCTGACGGGTCTTACGACTATCTCAACGGGCAACCG 420
QY 403 GGCAAGAACTTTAGACTAAATTTATATAGTTTCAAAATTAACAGAGTTAATGATTTGCCCAA 462
Db 421 GGCAAGACATCCGATCGCAGATGGTCAAGCCCTTCGACGCTTGGCTTGTATGTCGCTCC 480
QY 463 GACCAGCTGGCCATGATTTTCGCAAAATTTGTAGCTCTTGCATATTCAGGCTTTTAATC 522

Db 481 GAAAGCCTCGAGGTGATTCACCAAGTCACTCAGCATGCTACACACAGCCTCTTGTCTGCTC 540
QY 523 GACGATATAGAAGATAATGCTCCCTTGAGAAAGGGGACAGACCACTTCTCATTAACTTTC 582
Db 541 GACGAGCTGMAAGACACAGCGTCTCGCCCGGGCTTCCCTGTGCCCACTCCATCTTTC 600
QY 583 GGTGTACCTTCCATATATAACACCGCAAAATATATATGATTTTCAGAGCCATGCAACTTGT 642
Db 601 GGCATCCCCCAGACCATCAACACGTCACCACTAGTGTATTTCTACGCGCTGCAAGATTC 660
QY 643 TCGCAGCTTAACCAAAAGAGCCTTTGTATCATATTAATTTGATTTGATTTTCAACGAAGA 702
Db 661 CAAAAGCTCAAG-----ACCCCAAGCCGTCAGCATTTTCTCTGAAGAA 705
QY 703 TTGATCAATCTACATAGGGGACAAAGCTTGGATATATATCTGAGAGACTTTCTGCTGAA 762
Db 706 CTGCTCAACCTGCACCGGGCCAGGAATGGACCTCTTTTGGCGTGACAGCT-----C 759
QY 763 ATCATACCTACTCAGAGATGTTTGAATATGTTTGAATATGTTTGAATATGTTTGAATATGTTT 822
Db 760 ACCTGCCGACGAGGAGCGACTACCTAGAGATGTTTATCCCAACAGACAGCGCTGTGTC 819
QY 823 AGATTAACGTTGAGACTCATGGAAGCGCTGCTCTCTTCTCACACACGCGCATTCGTTG 882
Db 820 CCGTTGGGAATCAAGCTGATGCGAGCCGAGTCCGCGTCCGCGTGGACTGC----- 870
QY 883 GTTCTTTTCAAAATCTTCTGGGTATTTATCAGATTAGAGATGATTACTTGAATTTG 942
Db 871 GTCCGCTCGTCAACATCATCGACTGATCTTCCAGATCGCGCAGATTTATCATATCTG 930
QY 943 AAAGATTTCCAAATGTCAGGAAAGGCTTTGCTGAGCAATTCAGAGGGGAAAGTTA 1002
Db 931 TGGAACCGCGAGTACACGGCCAAACAAAGGCGATGTCCGAGGACCTCAGCGGCGAGTTC 990
QY 1003 TCTTTTCCATCGTCCACGCCCT 1025
Db 991 AGTTTCCCGTGTATCCACAGCAT 1013

RESULT 11

US-10-189-268-11
; Sequence 11, Application US/10189268
; Publication No. US20040005570A1
; GENERAL INFORMATION:
; APPLICANT: Nicholas M. Dean
; APPLICANT: C. Frank Bennett
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF GERANYLGERANYL DIPHOSPHATE SYNTHASE 1 EXP
; FILE REFERENCE: PUS-0021
; CURRENT APPLICATION NUMBER: US/10/189,268
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 131
; SEQ ID NO 11
; LENGTH: 51001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1331-1430, 34714-34813
; OTHER INFORMATION: n = A,T,C or G
US-10-189-268-11

Query Match 5.6%; Score 87.6; DB 16; Length 51001;
Best Local Similarity 52.2%; Pred. No. 1.4e-09;
Matches 288; Conservative 0; Mismatches 234; Indels 30; Gaps 3;

QY 475 ATAGTTTCGCAAAATGTTGAGCTCTTGCATTAATTCAGCCCTTTTAAATCGACGATATAGAA 534
Db 15383 ATTATTATTGAAGTGACAGAAATGTTGCAATATGCGAGTTTACTCATCGATGATATGAA 15442
QY 535 GATAATGCTCCCTTGAAGGGGACAGACCTTCTCACTTAATCTTCGGTGTACCTCC 594

Db 15443 GACAACTCAAACTCCGAGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCACT 15502
Qy 595 ACTATAACACCGCAATATATATGATTTTACAGAGCCATGCACTTGTATCGAGCTAAC 654
Db 15503 GTCACTAATTCGCCAATTAACGTATATTCCTTGGCTTGGAGAAAGTCTTAACCCCT 15558
Qy 655 ACAAGAAGCCCTTGTATCAATAATTTGATAGATTTTCAACGAAGAATTTGATCAATCTA 714
Db 15559 -----TGATCACCCAGATGAGTGAAGCTTTTACCGCCAGCTTTTGAACATC 15607
Qy 715 CATAGGGACAAGCTTGGATATATACTGGAGAGACTTTCTGCCTGAAATCATACCTACT 774
Db 15608 CATCAGGGACAAGCTTGAATATTTACTGAGGGATATTAACATTTG-----TCCACT 15661
Qy 775 CAGGAGATGATTTGATATGATTTGATGATTAATAAACAAGCGGCGCTTTTTCAGATTAACGTTG 834
Db 15662 GAAGAAGATATAAAGCTATGCTGCTGCAGAAAACAGGTGGAGTGTGTTGGATTAGCAGTA 15721
Qy 835 AGACTCATGGAAGCGCTGCTCTCTCCCTCACACCAAGCGCATTTGGTTGCTTCTTTCATA 894
Db 15722 GGTCTCATGAGTGTGTTCTCTGATTAACAAGAGA-----TTTAAACCGCTACTT 15772
Qy 895 AATCTTCTGGTATATTTATCAGATTAGAGATGATTACTGAAATTTGAAAGATTTCCAA 954
Db 15773 AATACACTTGGCTCTTTTCCAAATAGGATGATTATGCTAATCTACACTCCAAAGAA 15832
Qy 955 ATGTCCAGCGAAAAGCGTTTGTGAGGACATTAACAGAGGGAGTATCTTTTCCCATC 1014
Db 15833 TATAGTGAACAAAGATTTTGTGAAGATCTGACAGAGGAAAGTTCTCATTTCTACT 15892
Qy 1015 GTCCAGCGCCCTT 1026
Db 15893 ATTCATGCTATT 15904

RESULT 12
US-10-041-018-21
; Sequence 21, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080U31/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 2222
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-041-018-21

Query Match 5.4%; Score 84.2; DB 16; Length 2222;
Best Local Similarity 50.5%; Pred. No. 1.9e-09;
Matches 330; Conservative 0; Mismatches 293; Indels 30; Gaps 4;

Qy 374 AACCTTATAATCACATCTCTTTTGAACCTGGCAAGAACTTTAGACTAAATTAATAGTTC 433
Db 372 AGCCCTTTATACATACACAGAGATTTCTGCGACAGCAATTCGGCTCTGAGTTGGCTTGG 431
Qy 434 AAATTAACAGAGTTATGAAATTTGCCAAAGCCAGCTGGCCATAGTTTCGGAAATTTGTTG 493
Db 432 CTTTCAATCACTGGTTGCTCATACCGGGCGGAAAGTTGGCGCAGATCGAGACATTTGTC 491
Qy 494 AGCTCTGCATTAATTCAGCCCTTTTAAATCGACGATATAGAGATAATGCTCCCTTGAGAA 553
Db 492 AGATGCTGCATAATTCAGATTTGCTCATTTGATGATTAATGAAGACAAATTCGATCCCTGCA 551
Qy 554 GGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATAAACACCGCAAT 613

Db 552 GAGGTGTCCCGTGGCGCAATTCCTACTACGGGTGGCCAGCACCAATAAATGCGGCAACT 611
Qy 614 ATATCTATTTACAGAGCCATGCAACTTGTATCGCACTAAACACAAAAAGAGCCTTTGTATC 673
Db 612 ATGCACTCTTTCTGGCGCTGGAGAAGGTGACAGAGCTGGATCATCCGAGGCT----- 664
Qy 674 ATAAATTTGATTAACGATTTTCAACGAAGAAATGATCAATCTCATATAGGGGACAGGCTTGG 733
Db 665 -----ACCAAGGTGACACCGAAACAATTTGCTGAGCTGCACCGTGGACAGGGCATGG 716
Qy 734 ATATATATCTGGAGAGACTTTTCTGCTGCTGAAATCATATACCTACTCAGGAGATGATTTGAATA 793
Db 717 AGATCTATTTGGCGCCAGACGCTT-----CACGTGTCCAATCCAGTCCGATTAACAAGCTGA 770
Qy 794 TGGTTATGAATAAACAAGCGGCGCTTTTTCAGATTAACGTTGAGACTCATGGAAGCGCTGT 853
Db 771 TGACTGTGGGCAAACTGGCGGCTCTTTATGCTGGCCATTCGCTTATGCA--GCTGT 827
Qy 854 CTCCTTCTCTCACACACGCGCATTCGTTGGTTCTTTTCAATAATCTTCTGGGTATTTT 913
Db 828 TCAGCTTCCAAACAAGGAGGACTTATCGAAG-----TTGACGGCTATATTTGGGCTGTACT 881
Qy 914 ATCAGATTAAGATGATTTGAAATTTGAATTTGAAGATTTCCAAATGTCCAGCGAAAAGGCT 973
Db 882 TTCAATATAGCGAGCACTATTGCAATCTGAGTCTGAAAGAGTACACGGAGAACAGAGCT 941
Qy 974 TTGCTGAGCACATTAACAGAGGGGAAGTATCTTTTCCCATCGTCCACGCGCCTT 1026
Db 942 TGCCGAGGACTTGACGAGGAGGCAAGTTGCGGCTTCCCGGTAAATCCATGCGGTT 994

RESULT 13
US-10-001-192A-14
; Sequence 14, Application US/10001192A
; Publication No. US20040091958A1
; GENERAL INFORMATION:
; APPLICANT: Ooijen, Albert
; APPLICANT: Verdoes, Jan
; APPLICANT: Wery, Jan
; TITLE OF INVENTION: IMPROVED METHODS FOR TRANSFORMING
; TITLE OF INVENTION: PHAFFIA STRAINS, TRANSFORMED PHAFFIA STRAINS SO OBTAINED AND
; FILE REFERENCE: 24615-20104.01
; CURRENT APPLICATION NUMBER: US/10/001,192A
; CURRENT FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: EP 95203620.0
; PRIOR FILING DATE: 1995-12-22
; PRIOR APPLICATION NUMBER: EP96200943.7
; PRIOR FILING DATE: 1996-04-11
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82)....(1212)
US-10-001-192A-14

Query Match 4.5%; Score 70.6; DB 16; Length 1882;
Best Local Similarity 51.8%; Pred. No. 3.1e-06;
Matches 160; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 352 CAATATGAACCTTGATTTCAAAACCTTATAATCACATCTCTTTTGAACCTTGCAAGAAC 411
Db 130 CAGATGATATCGGCTCTTCTGAAACCTGATCATCTACCTAGAAAGAACCTTGGAAAGAA 189
Qy 412 TTTAGACTAAATTTAATAGTTTCAATTAACAGAGTATTAATTTGCCAAAAGACAGCTG 471
Db 190 ATTGATCACAACTCATCGAGGCTTTCAACTATTGTTGGATGTCAAGAGAGGATCTC 249

QY 472 GCATAGTTTCGCAAAATGTTAGCTCTTGCATAAATCCAGCCCTTTTAATCGCAGATATA 531
Db 250 GAGGTCTATCCAGAACGTTGTTGGCATGCTACATACCGCTAGCTTATTAATGACGATGTG 309
QY 532 GAAGATAATGCTCCCTTGAGAGGGGACAGACCACTTCTCATTAAATCTTCGGTGTACCC 591
Db 310 GAGGATTCATCGGTCTCTCAGGCGTGGGTGCGCTGTGCGCCATCTAAATTAACGGGATTCG 369
QY 592 TCCACTATATAACACCGCAAAATATATGTTATTTTCAGAGCCATGCAACTTGTATCGCAGCTA 651
Db 370 CAGACATATAACACTGCAAACTACGCTCTACTTTCTGGCTTATCAAGAGATCTTCAAGCTT 429
QY 652 ACCACAAA 660
Db 430 CGCCCAACA 438

RESULT 14

US-09-969-034-1324/c
; Sequence 1324, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poorima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1324
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 496..517, 531, 536
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-1324

Query Match 4.4%; Score 68.8; DB 11; Length 554;
Best Local Similarity 54.2%; Pred. No. 4.6e-06;
Matches 192; Conservative 0; Mismatches 147; Indels 15; Gaps 2;
QY 673 CATAAATTGATTACGATTTTCAACGAGAAATGATCAATACATACATAGGGGCAAGGCTTG 732
Db 422 CAAGATGCAGTGAAGCTTTTACCGCGCAGCTTTTGGAACTCCATCAGGGCAAGGCGCTA 363
QY 733 GATATATCTGGAGAGACTTTCGCTGCAATCATACCTACTCAGGAGATGTAATTGAAT 792
Db 362 GATATTACTGGAGGA-----TAATTACACTTGTCCACTGAAGAAGAAATATAAAGCT 309
QY 793 ATGTTATGAATAAAACAGCGGCGCTTTTCAGATTAAAGTTGAGACTCATGGAAGCGGTG 852
Db 308 ATGGTGTGCAAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTCTCATGCACTGTTTC 249
QY 853 TCTCCTTCTCACACCGGCGCATTCGTTGGTTCCTTTCATAAAATCTTCTGGGTATTATT 912
Db 248 TCTGATTACAAAGAAGATTAAACCGCTACTT-----AATACACTTGGGCTCTTT 198
QY 913 TATCAGATTACAGATGATTACTTGAATTTGAAAGATTTCCAAATGTCCAGCGAAAGGCG 972
Db 197 TTCCAAATTAGGATGATTGTGTAATCTACACTCCAAAGAAATGTAGTGAAACAAAAGT 138

QY 973 TTGCTGAGACATTAACAGGGGAAGTTATCTTTTCCCATCGTCCACGCCCTT 1026
Db 137 TTTTGTGAAGATGTGACAGAGGGGAAAGTTCTCATTTCTACTATTATGCTATT 84

RESULT 15

US-09-969-034-1924/c
; Sequence 1924, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poorima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1924
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 519, 560, 581, 585, 590, 600, 602, 628, 669, 670, 673, 679,
; LOCATION: 702
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-1924

Query Match 4.4%; Score 68.4; DB 11; Length 716;
Best Local Similarity 54.7%; Pred. No. 6.5e-06;
Matches 188; Conservative 0; Mismatches 141; Indels 15; Gaps 2;
QY 683 TTACGATTTTCAACGAGAAATGATCAATCTACATAGGGGACAGGCTTGGATATATACT 742
Db 409 TGAAGCTTTTACCGCGCAGCTTTTGGAACTCCATCAGGGCAAGGCGCTAGATATTTACT 350
QY 743 GGAGAGACTTTCTGCTGAAATCATACCTACTCAGGAGATCTATTGTAATATGGTTATGA 802
Db 349 GGAGGA-----TAATTACACTTGTCCACTGAGAAGAAATTAAGCTATGGTGCTGC 296
QY 803 ATAAAAACAGCGCGCTTTTCAGATTAAAGTTGAGACTCATGGAAGCGCTGTCTCTTCTCT 862
Db 295 AGAAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTCTCATGCACTGTTTCTCTGATTACA 236
QY 863 CACACCCAGCGCCATTCGTTGGTTCCTTTCATAAATCTCTCGGTATTATTTATCAGATTA 922
Db 235 AAGAAGATTAAACACCGCTACTT-----AATACACTTGGGCTCTTTTCCAAATTA 185
QY 923 GAGATGATTACTTGAATTTGAAAGATTTCCAAATGTCCAGCGAAAAAGGCTTTCGTGAGG 982
Db 184 GGGATGATTATGCTAACTACACTCCAAAGAAATAGTAGTGAACAAAGTTTGTGGAAG 125
QY 983 ACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCCACGCCCTT 1026
Db 124 ATCTGACAGAGGGGAAAGTCTCTCAATTCCTACTATTTCATGCTATT 81

Search completed: October 17, 2004, 08:40:44
Job time : 735.274 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:27:39 ; Search time 4877.9 Seconds
(without alignments)
11721.026 Million cell updates/sec

Title: US-10-041-018-1
Perfect score: 1569
Sequence: 1 aatattacatatagatag.....aatcgcgtaaatcgga 1569

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|----------|--------------------|
| 1 | 519.6 | 33.1 | 957 | 9 | CNS06GUS | AL398163 T7 end of |
| 2 | 510.8 | 32.6 | 990 | 9 | CNS06GEP | AL397607 T3 end of |
| 3 | 468.2 | 29.8 | 1021 | 9 | CNS06GIG | AL397742 T3 end of |
| 4 | 411.8 | 26.2 | 537 | 8 | AQ502594 | AQ502594 V52B7 mtn |
| 5 | 195 | 12.4 | 1115 | 9 | CNS06CKP | AL392639 T7 end of |
| 6 | 114 | 7.3 | 903 | 9 | AY407652 | AY407652 Homo sapi |
| 7 | 114 | 7.3 | 1076 | 5 | BX460174 | BX460174 BX460174 |
| 8 | 114 | 7.3 | 1417 | 3 | CR612465 | CR612465 full-leng |
| 9 | 114 | 7.3 | 2703 | 3 | CR603114 | CR603114 full-leng |
| 10 | 113.6 | 7.2 | 903 | 9 | AY407654 | AY407654 Mus muscu |
| 11 | 113.6 | 7.2 | 1341 | 3 | AK011980 | AK011980 Mus muscu |
| 12 | 103 | 6.6 | 815 | 4 | BI520491 | BI520491 603071690 |
| 13 | 99 | 6.3 | 905 | 5 | EX439729 | EX439729 EX439729 |
| 14 | 97.6 | 6.2 | 871 | 4 | BI560099 | BI560099 603253287 |
| 15 | 96.2 | 6.1 | 974 | 4 | EG675075 | EG675075 602621320 |
| 16 | 95.2 | 6.1 | 748 | 2 | BF610750 | BF610750 dd5h05.y |
| 17 | 95.2 | 6.1 | 1116 | 3 | AY222995 | AY222995 Schistos |
| 18 | 94.8 | 6.0 | 1077 | 4 | BG114452 | BG114452 602285755 |
| 19 | 94 | 6.0 | 514 | 1 | AA573392 | AA573392 nm53b07.s |
| 20 | 94 | 6.0 | 610 | 7 | CR360928 | CR360928 170005325 |
| 21 | 93.6 | 6.0 | 544 | 2 | BF400767 | BF400767 UI-R-CAO- |
| 22 | 93.6 | 6.0 | 546 | 2 | BE113864 | BE113864 UI-R-CAO- |
| 23 | 93.6 | 6.0 | 678 | 6 | CF146873 | CF146873 UI-HF-CB0 |
| 24 | 93.6 | 6.0 | 693 | 4 | BM721020 | BM721020 UI-E-B00- |

| | | | | | | |
|----|------|-----|------|---|----------|--------------------|
| 25 | 93.6 | 6.0 | 812 | 5 | EX846096 | EX846096 BX846096 |
| 26 | 92.8 | 5.9 | 803 | 9 | AY407653 | AY407653 Pan trogl |
| 27 | 92.8 | 5.9 | 873 | 6 | CD108157 | CD108157 AGENCOURT |
| 28 | 92.8 | 5.9 | 911 | 5 | BQ433460 | BQ433460 AGENCOURT |
| 29 | 92.2 | 5.9 | 698 | 7 | CN360927 | CN360927 170006000 |
| 30 | 91.2 | 5.8 | 924 | 5 | EX358978 | EX358978 BX358978 |
| 31 | 91 | 5.8 | 629 | 6 | CB272278 | CB272278 mai56e01. |
| 32 | 90.4 | 5.8 | 569 | 2 | AW270970 | AW270970 xs07a01.x |
| 33 | 89.6 | 5.7 | 797 | 5 | BU568661 | BU568661 AGENCOURT |
| 34 | 89.4 | 5.7 | 1007 | 5 | BQ886489 | BQ886489 AGENCOURT |
| 35 | 88.8 | 5.7 | 515 | 6 | CA880328 | CA880328 K0982B05- |
| 36 | 88.8 | 5.7 | 539 | 6 | CA870528 | CA870528 K0902G09- |
| 37 | 88.8 | 5.7 | 572 | 6 | CF198250 | CF198250 maj4ea06. |
| 38 | 88.8 | 5.7 | 777 | 6 | CB724192 | CB724192 UI-M-FY0- |
| 39 | 88.2 | 5.6 | 529 | 9 | CG619892 | CG619892 AGENCOURT |
| 40 | 88.2 | 5.6 | 807 | 7 | CK476319 | CK476319 AGENCOURT |
| 41 | 88 | 5.6 | 630 | 6 | CB190752 | CB190752 pl30e10.y |
| 42 | 88 | 5.6 | 907 | 5 | EX777578 | EX777578 BX777578 |
| 43 | 87.8 | 5.6 | 527 | 6 | CB296588 | CB296588 12B22037 |
| 44 | 87.8 | 5.6 | 859 | 5 | BQ938817 | BQ938817 AGENCOURT |
| 45 | 87.2 | 5.6 | 627 | 6 | CB215219 | CB215219 NISC_np03 |

ALIGNMENTS

CNS06GUS 957 bp DNA linear GSS 30-NOV-2001
T7 end of clone AS0AA007D07 of library AS0AA from strain CLIB 533
of Saccharomyces bayanus, genomic survey sequence.

AL398163 GI:12151423

GSS.

Saccharomyces bayanus

Saccharomyces bayanus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 957)

Souciat, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

20584711

11152876

2 (bases 1 to 957)

Bon, E., Neuvéglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,

Aigle, M. and Durrens, P.

Genomic exploration of the hemiascomycetous yeasts: 5.

Saccharomyces bayanus var. uvarum

FEBS Lett. 487 (1), 37-41 (2000)

20584715

11152880

3 (bases 1 to 957)

Genoscope.

Direct Submission

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of


```

/clone="AS0AA004A12"
/clone_lib="AS0AA"
/notes="end : T3"
<36..>670
/notes="similar to Saccharomyces cerevisiae ORF YPL069c [
BTS1 : geranylgeranyl diphosphate synthase ]
1 putative frameshift(s)"
/evidence=not_experimental
complement(<741..>989)
/notes="similar to Saccharomyces cerevisiae ORF YPL070w [
weak similarity to Vps9p ]"
/evidence=not_experimental

misc_feature
1470 AAATTTATAGCCTTATATATTTTCCACGATTTCTGAAATCCTTTTATCAGCACCGTT 1529
801 AAATTTGACGCTTGTATTTCTTCCATGACATTTTGATTTCTTCTCAATACCGCT 860

1530 AATGC-----TAGCGGTTACTGTCAAATCGCGGTAAATTCGCA 1569
861 AACACTGCAATGGCGCCCTACTGTCAAGTCACCTGTAAATTCACGA 906

RESULT 3
CNS06GIG
LOCUS
DEFINITION
T3 end of clone AS0AA004H03 of library AS0AA from strain CLIB 533
of Saccharomyces bayanus, genomic survey sequence.
ACCESSION
AL397742
VERSION
AL397742.1 GI:12150636
KEYWORDS
SOURCE
SACCHAROMYCES BAYANUS
ORGANISM
Saccharomyces bayanus
Saccharomycetes
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 1021)
Soudiet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekalia, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED
11152876
2 (bases 1 to 1021)
Bon, E., Neuveglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,
Aigle, M. and Durrens, P.
Genomic exploration of the hemiascomycetous yeasts: 5.
Saccharomyces bayanus var. uvarum
FEBS Lett. 487 (1), 37-41 (2000)
20584715
PUBMED
11152880
3 (bases 1 to 1021)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1..1021
/organism="Saccharomyces bayanus"
/mol_type="genomic DNA"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA004H03"
/clone_lib="AS0AA"
/notes="end : T3"
<1..>578
/notes="similar to Saccharomyces cerevisiae ORF YPL069c [
BTS1 : geranylgeranyl diphosphate synthase ]
1 putative frameshift(s)"
/evidence=not_experimental

misc_feature
1470 AAATTTATAGCCTTATATATTTTCCACGATTTCTGAAATCCTTTTATCAGCACCGTT 1529
801 AAATTTGACGCTTGTATTTCTTCCATGACATTTTGATTTCTTCTCAATACCGCT 860

1530 AATGC-----TAGCGGTTACTGTCAAATCGCGGTAAATTCGCA 1569
861 AACACTGCAATGGCGCCCTACTGTCAAGTCACCTGTAAATTCACGA 906

Query Match
Best Local Similarity 32.6%; Score 510.8; DB 9; Length 990;
Matches 717; Conservative 0; Mismatches 182; Indels 47; Gaps 5;

QY 631 ATGCAACTGTATCGAGCTAACCACAAAGAGCCTTTGTATCATATTTGATTACGATT 690
Db 1 ATGCAACTGTATCGCACTAATCTGC-AGACCGCTCTCTACGTAATAATTAATGACCATC 59

QY 691 TTCAACGAGAATTCATCAATCTACATAGGACAGAGCTTGTGATATATCTGATGATGATGAT 750
Db 60 TTATATGAGAAGTCAATCTATCATAGAGCCAGGTTTGGATATATTTGGAGAGAT 119

QY 751 TTCTGCTGAAATCATACCTACTCAGGAGATGTATTTGAATATGTTATGATGAATAAACA 810
Db 120 TTCTACCTGAAATCATACCTACTACCAAGAAATGTATTTGAATATGTTATGATGAATAAAG 179

QY 811 GCGCGCTTTTCAGATTACGTTGAGACTCATGGAAGCGCTCTCTCTCTCTCTCTCTCTC 870
Db 180 GTGTGCTCTTTTCAGATTACGCTTGAGGCTCATGGAAGCGTTCTCTCTCTCTCTCTCTC 239

QY 871 GCGCATTCCTGTTGCTCTTCATTAATCTCTCGGTATTTATTTATCAGATTAGATGAT 930
Db 240 GTCCATTCGCTGTTGCTCTTCATTAATCTTTTAGGTATGATTTATCAGATTAGAGACAT 299

QY 931 TACTGAAATTTGAAGATTTCCAAATGTCAGCGAAAGGCTTGTGCTGAGGACATTTACA 990
Db 300 TATCTAAATTTGAAGACTTCAAGATGTCATGAAAGGCTTGTGCTGAGGATATACCC 359

QY 991 GAGGGAGATTATCTTTTCCATCGTCCAGCGCCCTTAACTTCACTPAAACGAAAGGTC 1050
Db 360 GA-AGGAGCTCTCTTTTCCCATCATCTCAATGCTCTAAATTTCAACGAGCGCAAGCTCA 418

QY 1051 ACTGAGCAACAATGAAATTTCTAAGATTTCTCTGTTGAGGACAGATGATAAAGATATA 1110
Db 419 ACTGAGCAACAAGATGAAATTCATGAAATTTCTTTTAGGAGCTAATGATACAAAGTCTA 478

QY 1111 AAACCTAAGCTGATTCAAATCTGGAATTCGACACCAATTCATTTGGCTTACACCAAAAT 1170
Db 479 AAACCTGAAATTTGATGAAATTTCTAGAAATTCAGACCAATTCCTGAACTATACGAAGAT 538

QY 1171 TTATTAATCAATTTAGTGAATGATAAATAAATGATAAATAAGTATTTTACCTGAT 1230
Db 539 TTATTAATCAATTTAGTGAATGATAAATAAATGATAAATAAGTATTTTACCTGAT 598

QY 1231 TTGCTTTCGATTCGACACCGCCCAATTTTACATGACGAAATTTGTTATATATATAGAC 1290
Db 599 TTGACCCGCGATTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 658

QY 1291 CACTATCCGAATTTGAAATTAATTTGATCAATCAATCAATTTAGTGGAGAGATAGTACGAA 1350
Db 659 CATTATCCGAAGCTGAA-----TTAA 680

QY 1351 ATAAGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1409
Db 681 AAAAGCAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 740

QY 1410 TTGCATCATCTTTTGTATATATCTCAAAAGATCTCTTAGTTCGCAATAGTCAAAATCTTC 1469
Db 741 TTGCATCATCTTTTGTATATTTCAATAAATCTTTTAGTTCGCAATAGTCAAAATCTTC 800

```

LOCUS AO502594 537 bp DNA linear GSS 29-APR-1999
DEFINITION V5287 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', genomic survey sequence.
ACCESSION AO502594
VERSION AO502594.1 GI:4705140
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 537)
AUTHORS Roes-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
desRoches, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
Umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Iugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumaryale.yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTTCCTTCCTTGGAGATAC
Class: transposon-tagged.
Location/Qualifiers
1..537
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/lab_host="E. coli"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/notes="vector: PHSS6-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in PHSS6-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

misc_feature complement(<649..>939)
/note="similar to Saccharomyces cerevisiae ORF YPL070w [
weak similarity to Vps9p]"
/evidence=not_experimental
ORIGIN
Query Match 29.8%; Score 468.2; DB 9; Length 1021;
Best Local Similarity 76.1%; Pred. No. 7.3e-102;
Matches 649; Conservative 0; Mismatches 158; Indels 46; Gaps 4;
QY 724 CAAGCGCTGGATATACCTGGAGAGACTTCTGCTGAAATCATATACCTACTCAGAGATG 783
DB 1 CAAGGTTGGATATATTTGGAGAGATTTTCTACCTGAAATCATATACCTACCAAGAAATG 60
QY 784 TATTTGAATATGTTATGAATAAACAAGCGGCGCTTTTCAGATTAAGGTGAGACTCATG 843
DB 61 TATTTGAATATGTTATGAATAAACAAGCGGTGCTTTTCAGATTAAGGTGAGGTCATG 120
QY 844 GAAGCGCTGCTCTCTCCCTCACACGAGCGCATTCGTTGGTTCCTTTTCATAATCTTCG 903
DB 121 GAAAGCTTCTCTCTCTCTCACACGAGCGCATTCGTTGGTTCCTTTTCATAATCTTTA 180
QY 904 GGTATTTATTCAGATTAAGATGATTAATCTTGAATTTGAAAGATTTCCAAATGTCCAGC 963
DB 181 GGTATGATTTATCAGATTAAGATGATTAATCTTGAATTTGAAAGATTTCAAGATGTCCAAT 240
QY 964 GAAAGAGCTTTGCTCGAGGACATTAACAGAGGGGAAGTTATCTTTTCCCATTCGTCAGGC 1023
DB 241 GAAAGAGCTTCTCTCTCTCTCACCGA-AGGAAGCTCTCTTTCCCATCATCTCCATGCT 299
QY 1024 CTTAATCTCACTAAACGAAAGTCTAACTGACGACACAAATCTTAAGAAATCTC 1083
DB 300 CTAATTTTCAACCAAGCGCGAAGCTCAACTGACGACACGATGAATTCATGAATTCCT 359
QY 1084 CTCTTGAGGACAACTGATTAAGATATATAAACTAAAGCTGATTCAAATCTCGAATTCGAC 1143
DB 360 TTAATGAGGACTAATGATACAGCTTAACACTGAAATGATTAAGAAATCTTGAATTCAG 419
QY 1144 ACCAATTCATTTGCGCTACACAAATTTTATTAATCAATAGTGAATATGATAAAAT 1203
DB 420 ACCAATTCCTGAACTATACGAAAGATTTTATCAATCAGTTAGTGAACATGATAAAAT 479
QY 1204 GATAATGAATAATAGTATTTACCTGATTTGCTTCCGATTCGACACCGCACCAATTTA 1263
DB 480 GATGACAAAATTAAGTATTTGCGCGATTTGACCCGCTTCTTAATACCTCTACTAATTTA 539
QY 1264 CATGACGAATGTTATATAATAGACCACTTATCCGAATTTGCAATATAAATGATCAAT 1323
DB 540 CGTGACGAGCTATATATATAGTAGATCAITTTATCCGAAGTAA----- 584
QY 1324 CAAATTAAGTGAGGAGATAGTACGAATAAAGCTTCTCTCTCTCTCTCTCTCTCTCTAT 1383
DB 585 -----TTAAAAAAGCAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 621
QY 1384 ACATAC-GATTTATATATAGTTTATGATCATCTTTTGTATATATATCAAAAGATC 1442
DB 622 ATATACAGTATTTATATAGTATATTTGATCATCTTTTGTATATTTCAATAATC 681
QY 1443 TCTTAGTTGCAATATAGTCAAACTCTTCAAAATTTATAGCTTTATTTTTCACGATTT 1502
DB 682 TTTTAGTTGCAATATAGTCAAACTCTTCAAAATTTGATTTTTCACGATTT 741
QY 1503 CTGAAATCTCTTTTATCAGACCGTTAATGC-----TAGCGGTTACTGTCAATCGC 1556
DB 742 TTGATTTCTCTCTCTCAATACGCTAACCTTGATTTGGCGCTACTGTCAAGTACC 801
QY 1557 GGTAAATTCGGA 1569
DB 802 TGTAAATTCAGA 814
QY
DB
RESULT 4
AO502594/c

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 903)
 Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarizwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 14671302

2 (bases 1 to 903)
 Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarizwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 This sequence was made by sequencing genomic exons and ordering
 them based on alignment.

FEATURES
 source
 1..903
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1..>903
 /gene="GGPS1"
 /locus_tag="HCM2962"

ORIGIN
 Query Match 7.3%; Score 114; DB 9; Length 903;
 Best Local Similarity 52.1%; Pred. No. 1.7e-16;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAGCTTGATTTCAAACCTTATAATCACATCCTTTTGAACCTGCGCAAGACTTTAGAC 418
 |||||
 Db 26 AAGAATTTCTTAGAACCTTATAATCTTCTCAGTTTACCAGGTAAACAGTGAGAA 85
 |||||

QY 419 TAAATTTAATAGTTCAAATTAACAGAGTTATGATTTTGCCAAAGACCGCTGGCCATAG 478
 |||||
 Db 86 CCAAACTTTCACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGCAAGCTACAGATTA 145
 |||||

QY 479 TTTCGCAATTTGTGAGCTCTCCATAATTTCCAGCTCTTTTAACTCGAGATATAGAATA 538
 |||||

Db 146 TTATGAGTGAACAGAAAGTTGTCATATATCCAGTTTACTCATCGATGATTTGAAGACA 205
 |||||

QY 539 ATGCTCCCTTGAGAAGGGGACAGACCACCTTCTCACTTAATCTTCGTTGTAACCTCCACTA 598
 |||||

Db 206 ACTCAAACCTCCGACGTGGCTTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 265
 |||||

QY 599 TAAACACCGCAATATATGTTTTCAGAGCCATGCACTGTGTATCGCAGCTAACCAACA 658
 |||||

Db 266 TCAATTCGCAATTTACGTGTTTTCCTGGCTGGAGAAAGTCT----- 310
 |||||

QY 659 AAGAGCCTTTGTATCAATAATTTGATTTAGATTTTCAACGAAGATTTGATCAATCTACATA 718
 |||||

Db 311 TAACCTTGATACCCAGATGCACTGAAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 370
 |||||

QY 719 GGGGACAAGGCTGGATATATCTACTGAGAGACTTTCTGCTCGTGAATCATACCTACTCAGG 778
 |||||

Db 371 AGGACAAGGCTAGATATTTACTTGGAGGA-----TAAATTACATCTTGTCCTCCACTGAAG 424
 |||||

QY 779 AGATGTTATGATATGTTATGATTAATAACAGCGCGCTTTTCAGATTAACGTTGAGAC 838
 |||||

Db 425 AAGAATATAAAGCTATGCTGTCAGAAAAACAGGTGGAGCTTTTGGATTTAGAGTAGTGC 484
 |||||

QY 839 TCATGGAAGCGCTGCTCTCTCTCTCACACCAACGCGCAATTCGTTGGTTCTTTTCAATAATC 898
 |||||

Db 485 TCATGCAAGTGTCTCTGATTTACAAAGAGATTTTAAACCGCTACTT-----AATA 535
 |||||

QY 899 TTCTGGGTATTTATATACAGATTAAGATTAATCTTGAATTTGAAGATTTCCAAATGT 958
 |||||

Db 536 CACTTGGCTCTTTTCCAAATTTAGGATGATTTATGCTAATCTACACTCCAAAGATATA 595
 |||||

QY 959 CCAGCGAAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTATCTTTTCCATCGTCC 1018
 |||||

Db 596 GTGAAAAACAAGTTTTGTGAAGATCTGACAGAGGGAAGTCTCTCAATTCCTACTATTC 655
 |||||

QY 1019 AGGCCCTT 1026
 |||||

Db 656 ATGCTATT 663
 |||||

RESULT 7
 EX460174 1076 bp mRNA linear EST 06-MAY-2004
 LOCUS BX460174 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
 DEFINITION CS0DF011YP23 5-PRIME, mRNA sequence.
 ACCESSION EX460174
 VERSION EX460174.2 GI:47072855
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1076)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 22, 2003 this sequence version replaced gi:31037026.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to sequence cluster 868.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0DF011CH12QF1&c=868.f.

FEATURES
 Location/Qualifiers
 1..1076
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DF011YP23"
 /tissue_type="FETAL BRAIN"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL BRAIN"
 /notes="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
 vector. Library was not normalized."

ORIGIN
 Query Match 7.3%; Score 114; DB 5; Length 1076;
 Best Local Similarity 52.1%; Pred. No. 1.7e-16;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAGCTTGATTTCAAACCTTATAATCACATCCTTTTGAACCTGCGCAAGACTTTAGAC 418
 |||||

Db 251 AAGAATTTCTTAGAACCTTATAATCTTCTCAGTTTACCAGGTAAACAGTGAGAA 310
 |||||

QY 419 TAAATTTAATAGTTCAAATTTAACAGAGTTATGATTTTGCCCAAGACCGCTGGCCATAG 478
 |||||

Db 311 CCAAACTTTCACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGCAAGCTACAGATTA 370
 |||||

QY 479 TTTCGCAATTTGTGAGCTCTTGCATTAATTCAGCCCTTTTAACTCGAGATATAGAATA 538
 |||||

Db 371 TTATGAGTGAACAGAAATGTTCATAATGCGAGTTTACTCATCGATGATTTGAAGACA 430
 |||||

QY 539 ATGCTCCCTTGAGAAGGGGACAGACCACCTTCTCACTTAATCTTCGTTGTAACCTCCACTA 598
 |||||

| ORIGIN | Query Match | Best Local Similarity | 7.3%; Score 114; DB 3; Length 1417; |
|------------|--|--|-------------------------------------|
| | Matches 348; Conservative | 52.1%; Pred. No. 1.8e-16; | Mismatches 290; Indels 30; Gaps 3; |
| QY | 359 | AAAGCTTGGATTCAAAACCTTATAATACATCATCTTTTGAACCTCGCAAGAACTTTAGAC | 418 |
| Db | 249 | AAAGAAATCTCTTAGAACCCCTATAAATCTTCTCAGTTTACCAGGTAAACAAGTCAGAA | 308 |
| QY | 419 | TAAATTTAAATAGTTCAAAATTAACAGAGATTATGAATTTGCCAAAGACCGCTCGCCATAG | 478 |
| Db | 309 | CCAACTTTTACAGGCGATTATATCATTTGGCTGAAAGTTCCAGAGGACAAAGCTACAGATTA | 368 |
| QY | 479 | TTTCGCAATTTGTTGAGCTCTTCGATAAATTCAGCCCTTTTAAATCGACATATAGAATA | 538 |
| Db | 369 | TTATTGAAGTGCAGAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATATTGAAGACA | 428 |
| QY | 539 | ATGCTCCCTTGAAGAGGGGACAGACCACCTTCTCATTATCTTCGTTGTTACCTCCACTA | 598 |
| Db | 429 | ACTCAAACTCCGACGTGGCTTTCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA | 488 |
| QY | 599 | TAAACACCGCAATTAATATATATTTTTCAGAGCCATGCACTTGTATCGAGCTAAACCAA | 658 |
| Db | 489 | TCAATTTCTGCAATTAAGTGTATTTCTCTTGGCTTGGAGAAAGTCT | 533 |
| QY | 659 | AGAGGCTTTGTATCATATTTTGTATACGATTTTTCACGAAATGATCAATCTACATA | 718 |
| Db | 534 | TAACTCTGATCCCGAGATGAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC | 593 |
| QY | 719 | GGGGACAAGGCTTTGGATATATATCTCTGAGAGACTTTCTGCTGCAATCATACCTACTCAG | 778 |
| Db | 594 | AGGACACAGGCTTAGATATTTACTTGGAGGGA | 647 |
| QY | 779 | AGATGATTTGAATATGTTATGCAATAAAACAGGGGGCTTTTCAGATTAAAGTTTCCAAATGT | 958 |
| Db | 648 | AAAGAATATAAGACTATGTTGCTGCAGAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC | 707 |
| QY | 839 | TCATGAGAGCGTGTCTCTCTTCTCACACCGGCAATTCGTTGGTTCTTTTCATAAATC | 898 |
| Db | 708 | TCATGAGATGTTTCTCTGATTAACAAGAAGATTTAAACCCGCTACTT | 758 |
| QY | 899 | TTCTGGGTATTTATCAGATTAGAGATGATTCTTGAATTTTGAAGATTTTCCAAATGT | 958 |
| Db | 759 | CACTTGGGCTCTTTTCCAAATAGGATGATATGCTAACTACACATCCCAAGATATA | 818 |
| QY | 959 | CCAGCGAAAAGGCTTTGCTGAGGACATTCAGAGGGGAAGTATCTTTTCCCATCGTCC | 1018 |
| Db | 819 | GTGAAAAACAAGTTTTTGTGAAGATCTGACAGAGGGAAGTTCTCAITTCCTACTATTC | 878 |
| QY | 1019 | ACGCCCTT 1026 | |
| Db | 879 | ATGCTATT 886 | |
| RESULT 9 | | | |
| LOCUS | CR603114 | | |
| DEFINITION | full-length cDNA clone CS0DJ001YK22 of T cells (Jurkat cell line) | | |
| ACCESSION | CR603114 | | |
| VERSION | CR603114.1 | GI:50483921 | |
| KEYWORDS | HTC; CNSLT cdNA. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| TITLE | Li, W.B., Gruber, C., Jessee, J. and Polayes, D. | | |
| JOURNAL | Full-length cDNA libraries and normalization | | |
| REMARK | Unpublished | | |
| CONTACT | Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue | | |
| REFERENCE | Genoscope. | | |
| AUTHORS | Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : | | |
| TITLE | BP 191 9106 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr | | |
| JOURNAL | - Web : www.genoscope.cns.fr) | | |
| COMMENT | 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1. .1417 | | |
| | /organism="Homo sapiens" | | |
| | /mol_type="mRNA" | | |
| | /db_xref="taxon:9606" | | |
| | /clone="CS0DJ001YK22" | | |
| | /tissue_type="Placenta Cot 25-normalized" | | |
| | /plasmid="pCMVSPORT_6" | | |

Paraday Avenue
2 (bases 1 to 2703)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
Location/Qualifiers
1..2703
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ001YK22"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 7.3%; Score 114; DB 3; Length 2703;
Best Local Similarity 52.1%; Pred. No. 2e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
QY 359 AAGCTTGATTTCAAAACCTTATATACATCCTTTTGAACCTGGCAAGACTTTAGAC 418
Db 1533 AAGAAGTCTCTAGAACCTTATAATCTTACTTCCAGTTACCGTTAAACAAGTGAGAA 1592
QY 419 TAAATTTAATAGTTCAATTTACAGAGTTATGAATTTGCCAAAGACCACTGGCCATAG 478
Db 1593 CCAAACTTTACAGGCAATTAATCATTTGGCTGAAGTTCCAGAGCAAGCTACAGATTA 1652
QY 479 TTTGCAAAATTTGTGAGCTCTTGCATAATTTCCAGCCTTTTAAATCGACATATAGAGATA 538
Db 1653 TTATTCAGTGACAGAAATTTGCAATATGCCAGTTTACTCATCGATGATATTGAAGACA 1712
QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTCTCAGTATCTTCGTTCCGTTACCTCCACTA 598
Db 1713 ACTCAAACTCCGACCTGGCTTTCCAGTGGCCACAGCAATCTATGGAATCCCATCTGTCA 1772
QY 599 TAAACACCCCAATATATATATATTTTCCAGAGCCATGCACTTTGATCGCAGTAAACACAA 658
Db 1773 TCAATTTCTGCCAATACGTTATTTCTTGGCTTGGAGAAAGTCT----- 1817
QY 659 AAGAGCCTTTGTATCATATTTGATTTACGATTTTCAACGAGAAATTTGATCAATCTACATA 718
Db 1818 TAACCTTTGATCACCAGATGCAATGAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 1877
QY 719 GGGGCAAGGCTTGGATATATCTTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGG 778
Db 1878 AGGGAACAAGCCCTAGATATTTACTTGGAGGA-----TAATACATCTGTCCCACTGAAG 1931
QY 779 AGATGTTATTTGAATATGTTTATCAATAAACAAGCGGCGCTTTTCAGATTAACCTTCAGAC 838
Db 1932 AAGAATTAAGATATGTTGCTGCAAGAAACAGGTGGACTGTTTGGATAGCAGTAGGTC 1991
QY 839 TCATGAAGCGCTGTCT 898
Db 1992 TCATGAGTGTCT 2042
QY 899 TTTCTGGGTATTTATTCAGATTTAGATGATTTACTTGAATTTGAAGATTTTCCAAATGT 958
Db 2043 CACTTGGGCTCTTTTCCAAATTTAGGAGATTTATGCTATCTACCTCCAAAGATATA 2102
QY 959 CCAGCAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAGTATCTTTTCCCATCTCTCC 1018
Db 2103 GTGAACAACAAGTTTTTGTGAGATCTGACAGAGGGAAGTTCTCTCAATCTCTACTATTC 2162
QY 1019 ACGCCCTT 1026
| | | | |

Db 2163 ATGCTATT 2170
RESULT 10
AY407654 903 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus GBS1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION Genomic survey sequence.
ACCESSION AY407654
VERSION AY407654.1 GI:39763625
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 903)
REFERENCE
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 903)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..903
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>903
/gene="GSP1"
/locus_tag="HCM2962"
ORIGIN
Query Match 7.2%; Score 113.6; DB 9; Length 903;
Best Local Similarity 51.9%; Pred. No. 2.1e-16;
Matches 350; Conservative 0; Mismatches 294; Indels 30; Gaps 3;
QY 353 AAAATGAAGCTTGATTTCAAAACCTTATAATCACATCTCTTTTGAACCTGGCAAGACT 412
Db 20 AAGCTGAGAGGATCTCTTAGAGCCCTTAGAGCTTACTTCTAGTTACCAAGTTAAACAGG 79
QY 413 TTAGACTTAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACCACTGG 472
Db 80 TGGAAGCAAACTTTCCAGGCAATTTAACTACCTGGCTGAAAGTTCCAGAGACAAGCTAC 139
QY 473 CCATAGTTTTCGAAAATTTGTGAGCTCTTGCAATTTCCAGCCTTTTAAATCGACATATAG 532
Db 140 AGATTATCATTTGAAGTGACTGAAATGTGCATAATGCCAGTTTACTCATTTGATGATATG 199
QY 533 AAGATTAATGCTCCCTTTGAGAGGGGACAGACCACTTCTCATTTAATCTTCGGTGTACCT 592
Db 200 AAGACAGTTTCAAAGCTCCGACGTTTCCAGTGGCTCACAGATCTATGTTGTCCCAT 259
QY 593 CCACATAACACCGCAATTTATATGTTTACAGAGCCATGCAACTTTGATATCGACGTAA 652
Db 260 CTGTCTATTAATCTGCCAATTTACGTCTACTTCTTGGACTGGAAAAAGTCT----- 310
QY 653 CCACAAAGAGCCTTTGTATCATTAATTTGATTTAGGATTTTCAACGAGAGATTTGATCATC 712
Db 311 -----TAACCTTTGATCACCCGATGCGGTGAAGCTTTTCAACGCGACGTTCTTGAAC 364
QY 713 TACATAGGGGCAAGGCTTGGATATATCTGAGAGACTTTTCTGCTGAAATCATACCTA 772

Db 365 TTCAACAGGACAGCCCTCGATATTTACTGGGAC-----ACCTACACTGTGCCAA 418
QY 773 CTCAGGAGATGATTTGAATATGTTTGAATATAAAGACAGCGCCCTTTTCAGATTAACT 832
Db 419 CTGAAGAAGATATAAAGCCATGGTGTTCGAGAAGACAGGTGTTTGTGTTGATTAGCAG 478
QY 833 TGAGACTCAGTGAAGCGCTGTCTCTCCCTTCACACACCGCCCATTCGTTGGTTCCTTCA 892
Db 479 TAGGCTTATGAGCTGTCTCTGATTACAAAGAAGATCTAAAGCACTGCCTTG----- 532
QY 893 TAAATCTTCTGGTATTTATTCAGATTAGATGATTAATCTTGAATTGAAGATTTC 952
Db 533 ---ACACACTTGGCTCTTTTCCAGATTAGAGATGATTATGCCAATCTACACTCCAAAG 589
QY 953 AAATGTCACGCAAAAGGCTTGTCTGAGCATTACAGAGGGAAGTATCTTTTCCCA 1012
Db 590 AATACAGTGAACAAAGTTCTGTGAAGCTTGACAGAGGAAGTCTCATTTCCCA 649
QY 1013 TCGTCCAGCCCTT 1026
Db 650 CTATCCATGCCATT 663

RESULT 11

AK011980
LOCUS DEFINITION Mus musculus 1341 bp mRNA linear HTC 03-APR-2004
enriched library, clone:2610304H04 product:geranylgeranyl
diphosphate synthase 1, full insert sequence.
ACCESSION AK011980.1 GI:12848439
VERSION HTC; CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi.N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE 11076861
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation

JOURNAL
REFERENCE
AUTHORS

of 60, 770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1341)
Arai,K., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hirao,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasakawa,T., Kato,H., Kawai,J., Kojima,Y., Komano,H., Kouda,M.,
Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saifu,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,K., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

TITLE
JOURNAL

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCGAGTATTAATTAATCCCTCCCTCCCTCCCT 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES
source

Location/Qualifiers
1..1341
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM.DB:2610304H04"
/clones="2610304H04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
365..1075
/note="unnamed protein product; geranylgeranyl diphosphate
synthase 1 (MGD)MG1:1341724, GB|NM_010282, evidence:
BLASTN, 100%, match=712)
putative"
/codon_start=1
/protein_id="BA827955.1"
/db_xref="GI:12848440"
/translation="MEKTEKAEKILPEYRLLQPGKQVSKLSQAFNHLKVPED
KQIITIEVTEMLHNSLIDIEDSKLRGFPVAHSIYGVPSVINSANVYVFLGLEK
VLTLPDPAVKLTROLLLEHOGQGLDIYWRDVTCTPTREYKAMVLQKGGFLGLAV
GLMOLPSDYKEDLKLPLDGLTGLFFQIRDDYANLHSEYSENKSFCHDLTEGKFSFTI
HAWSRPESTQVSLNQF"

CDS

ORIGIN

Query Match 7.2%; Score 113.6; DB 3; Length 1341;
Best Local Similarity 51.9%; Pred. No. 2.2e-16;
Matches 350; Conservative 0; Mismatches 294; Indels 30; Gaps 3;
QY 353 AAAATGAAGCTTGATTCAAAACCTTATAATCACCCTCTTTTGAACCTGGCAAGAACT 412
Db 384 AAGCTGAGAGGATCTCTCTAGAGCCCTATAGGTACTTACTTACAGGTAAACAGG 443
QY 413 TTAGACTAATTAATAGTTCATAATTAACAGAGTTATGATTTGCCCAAGACCACTGG 472

Db 444 TGAGAAGCAAACTTTCACAGSCATTTATCACTGGCTGAAAGTTCCAGAGACAAGCTAC 503
Qy 473 CCATAGTTTCGCAAAATGTTGAGCTCTTGCATAATATCCAGCTTTTATTCACGATATAG 532
Db 504 AGATTATCATTTGAAGTGAATGTTGCATAATATCCAGCTTTTACTCATTTGATGATATG 563
Qy 533 AAGTAATATGCTCCCTTGGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCT 592
Db 564 AAGACAGTTCAAAGCTCCGAGCTGGTTTCCAGTGGCTCAGAGCACTATGTTGGTCCCAT 623
Qy 593 CCACATATAACACCCCAATATATATATTTTCCAGAGCCATGCAACTTGTATGCGAGTAA 652
Db 624 CTGTCAATTAATCTGCCAATTAAGTCTTACTTCTTGGAGTGAAGAGTCT----- 674
Qy 653 CCACAAAAGAGCCTTTGTATCAATTAATTTGATTAACGATTTTCAACGAAGATGATCAATC 712
Db 675 -----TAACCTTGTATCACCAGGATGCGGTGAAGCTTTTCAACGCCAGCTTCTGGAAC 728
Qy 713 TACATAGGGGACAAAGCTTGGATATATATCTGAGAGACCTTTCTGCTGAAATCATACCTA 772
Db 729 TTCACTAGGAGCAAGGCTTCGATATTTTCTGAGGAGC-----ACCTACACTTGTCCA 782
Qy 773 CTCAGAGATGATTTGAATATGTTTATGAATAAACAAGCGGCGCTTTTCAGATTAACGT 832
Db 783 CTGAAGAAGATATAAAGCCATGTTGTTGAGAGACAGTGGTGTGTTGGATAGCAG 842
Qy 833 TGAGACTGAGAGCGCTGCTCTCTCTCTCAACACGCGCATTTGTTGGTTCCTTTCA 892
Db 843 TAGGTCTTATGACGCTGTTCTCTGATTAACAAAGAGATCTAAAGCCACCTGCTTG----- 896
Qy 893 TAAATCTTCGGGTATTTATATACATTAAGATGATTTGATTTGAAAGATTTTCC 952
Db 897 ---ACACACTTGGCTCTTTTCCAGATTAGAGATGATTTGCAATCTACACTCAAAG 953
Qy 953 AAATGTCCAGCAAAAAGCTTTGCTGAGGACATTAACAGAGGAGATTTATCTTTTCCA 1012
Db 954 AATACAGTGAACAAGTTCTGTGAAGCTTTGACAGAGGGAAGTTCTCATTTCCCA 1013
Qy 1013 TCGTCCAGCCCTT 1026
Db 1014 CTATCCATGCCATT 1027

RESULT 12
BI520491
LOCUS 603071690F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5163599 5',
DEFINITION mRNA sequence.
ACCESSION BI520491 GI:15345283
VERSION BI520491.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 815)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NIH-MGC <http://mgc.nci.nih.gov/>.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M11406 row: c column: 24
High quality sequence stop: 813.
Location/Qualifiers
1..815
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5163599"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dr primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."

ORIGIN

Query Match 6.6%; Score 103; DB 4; Length 815;
Best Local Similarity 52.0%; Pred. No. 7.5e-14; Indels 31; Gaps 4;
Matches 348; Conservative 0; Mismatches 290;
Qy 359 AAAGCTTGATTTCAAAACCTTATAATCACATCCTTTTGAACCTGCGCAAGAACTTTAGAC 418
Db 113 AAAGAATTTCTTAGAACCTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 172
Qy 419 TAAATTTAATAGTTCAAAATTAACAGATTAATGAATTTGCCAAAGACAGCTGGCCATAG 478
Db 173 CCAAACTTTACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACAGCTACAGATTA 232
Qy 479 TTTCCGAAATTTGAGCTCTTGCATAATTTCCAGCTTTTAAATCAGCATATATAGAATA 538
Db 233 TTTATGAAGTGACAGAAATGTTGCTAATGCTGATGCTGATGCTGATGCTGATGCTGATG 292
Qy 539 ATGCTCCCTTGAAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACATA 598
Db 293 ACTCAAACTCCGACGCTGCTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 352
Qy 599 TAAACAACCCAAATTAATATATATTTTTCAGAGCAATGCAACTTGTATCGAGCTAACACAA 658
Db 353 TCAATTTCTGCAATTAACGTGTATTTTCTTGGCTTGGAGAAAGTCT----- 397
Qy 659 AAGAGCTTTGTATCATATAATTTGATTAACGATTTTCAACGAAGAAATGATCAATCTACATA 718
Db 398 TAACTCTTGAATCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 457
Qy 719 GGGGCAACAGCTTGGATATATATCTGAGAGACTTTCTGCTGAAATCATACCTACTCAGG 778
Db 458 AGGGAACAAGGCTTAGATATTTACTGAGGGA-----TAATTAACACTTGTCCCACTGAAG 511
Qy 779 AGATGATTTGAATATGTTTATGAATAAACAAGCGGCGCTTTTTCAGATTAACGTTGAGAC 838
Db 512 AAGATATAAAGCTATGTTGCTGCTGAGAAACAGGTTGGTGGATTAAGCTAGTAGTTC 571
Qy 839 TCATGGAAGCGCTGCTCTCTCTCTCCACACAGCGCATTCGTTGGTTCCTTTTCATATAATC 898
Db 572 TCATGAGTGTGTTCTCTGATTAACAAGAGATTTAAACCGCTACTT-----AATA 622
Qy 899 TTTCTGGTATTTATTTATTCAGATTTAGAGATGATTTGAATTTGAAGATTTCCAAATGT 958
Db 623 CACTTGGGCTCTTTTCCAAATTTAGGATGATTTATGCTAATCTACACTCCAAAGAAATATA 682
Qy 959 CCAGCGAAAAGGCTTGTGCTGAGGACATTTACAGAGGGG-GAAGTTATCTTTTCCCATCGTC 1017
Db 683 GTGAAAAACAAAAGTTTGTGAAGATCTGACAGAGGGGCAAGTTCTCATTTCTACTATT 742
Qy 1018 CAGCCCTT 1026
Db 743 CATGCTATT 751

RESULT 13
BX439729
LOCUS

BX439729 905 bp mRNA linear EST 04-MAY-2004

DEFINITION BX439729 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE014Y115
 5-PRIME, mRNA sequence.
 ACCESSION BX439729
 VERSION BX439729.2 GI:47012615
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 905)
 Li W.B., Gruber C., Jessee J. and Polayes D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 15, 2003 this sequence version replaced gi:30785818.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to sequence cluster 868.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?S=CS0DE014AE08QP1&c=868.f.
 Location/Qualifiers
 1. 905

FEATURES

source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DE014Y115"
 /tissue_type="PLACENTA"
 /clone_lib="Homo sapiens PLACENTA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 6.3%; Score 99; DB 5; Length 905;
 Best Local Similarity 51.8%; Pred. No. 7e-13;
 Matches 338; Conservative 2; Mismatches 282; Indels 31; Gaps 4;

QY 359 AAAGCTTGATTCAAACCTTATAATACAGAGTTATGAATTTGCCCAAGACGAGCTGGCCATAG 418
 |||||
 Db 209 AAAGAATCTCTAGAACCTTATAAATCTTACTTCAGTTACCGAGGTAAACAAAGTGAGAA 268
 QY 419 TAAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACGAGCTGGCCATAG 478
 |||||
 Db 269 CCAAACTTTCACAGGCATTTAATCATTTGGCTGAAAGTTCAGAGGACAAGCTACAGATTA 328
 QY 479 TTTTCGCAATTTGTAGCTCTTGATTAATTCAGACCTTTTAAATCGACGATATAGAAGATA 538
 |||||
 Db 329 TTAATGAAGTGACAGAAATGTGCATATGCGCAGTTTACTCATCGATGATATTGAAGACA 388
 QY 539 ATGCTCCCTTTGAGAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 598
 |||||
 Db 389 ACTCAAACTCCGACGTGGCTTCCAGTGCCACAGCATCTATGGAATCCCATCTGTCTCA 448
 QY 599 TAAACACCGCAATTAATGATTAATTCAGAGCCATGCAACTTGTATCGCAGCTAACCAACA 658
 |||||
 Db 449 TCAATTCGCCAATTAGTGTGTTTCTTGGCTTGGGAAGTCT----- 493
 QY 659 AAGAGCTTTGTATCATTAATTTGATACGATTTTCAACGAAGAATTGATCAATCTACATA 718
 |||||
 Db 494 TAAACCTTGTATCACCAGATGCGATGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 553
 QY 719 GGGGACAGGCTTGGATATATCTGAGGAGACTTTCTGCTCGTAATCATACCTTACTCAGG 778
 |||||
 Db 554 AGGGACAGGCTTAGATATTTACTGGAGGA-----TAATTACATTTGCCACTGAAG 607

QY 779 AGATCTATTGTAATGTTATGATATAAAACAGGGGGCCCTTTTCAGATTAAAGTTGAGAC 838
 |||||
 Db 608 AAGAATATAAAGCTATGTTGCTGCGAGAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC 667
 QY 839 TCATGGAAGCGCTGCTCTCTCTCCACACACCGCCATTCGTTGGTTCCTTTTCATATAATC 898
 |||||
 Db 668 TCATCAGCTGTTCTCTGATTACAAAGAGATTTAAACCGCTACTT-----AATA 718
 QY 899 TTCTGGGTATTTATCAGATTAGAGATGATTAATCTGAATTTGAAAGATTTCCTCAATGT 958
 |||||
 Db 719 CACTGGGCTCTTTTCCAAATTAGGATGATTAATCTACATCCCAAGAAATAT 777
 QY 959 CCAGCGRAAAGGCTTCTGAGGACATTCAGAGGGGAAGTATCTTTTCC 1011
 |||||
 Db 778 AGTGAACAAAAGTTTTTGTGAAGATCTGCACAGAGGAAGTCTCATTTTCYWC 830

RESULT 14

BI560099

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 871)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHAM1748 row: m column: 21

High quality sequence stop: 781.

Location/Qualifiers

1. 871

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5295932"

/lab_host="DH10B"

/clone_lib="NIH MGC 97"

/notes="Organ: testis; Vector: pBluescriptR (modified

pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI

(gtcgag); Oligo-dT primed using primer

5'-TTTTTTTTTTTTTN-3', size-selected for average

insert size 2.2 kb and normalized to ROT 5. This is a

primary library enriched for full-length clones and

constructed using the Cap-trapper method (Carninci, in

preparation). Library constructed by M. Brownstein

(NIH/NHGRI, National Institutes of Health). Note: this is

a NIH_MGC Library."

ORIGIN

Query Match 6.2%; Score 97.6; DB 4; Length 871;

Best Local Similarity 53.0%; Pred. No. 1.5e-12;

Matches 269; Conservative 0; Mismatches 219; Indels 20; Gaps 2;

QY 359 AAAGCTTGATTCAAACCTTATAATACATCTCTTTTGAACCTGCGAAGACTTAGAC 418

|||

Db 203 AAAGAATCTCTTAGAACCTTATAATATTACTTACGTTACCAGGTAAACAGTAGAA 262

```
QY 419 TAAATTTAATAGTTCAATTAACAGAGTTATGAAATTTGCCAAAGACAGAGCTGGCCATAG 478
D 263 CCAAACTTTCACAGGCAATTAATCATGTCGTAAGAGTTCCAGAGACAGCTACAGATTA 322
QY 479 TTTCGCAATTTGTTGAGCTCTTTGCATAAATCCAGCCTTTTAAATCGACGATATAGAAGATA 538
D 323 TTATTTGAAGTACAGAAATGTTGCATATGCGAGTTTACTCATCGATGATATGAAGACA 382
QY 539 ATGCTCCCTTGAGAAAGGACAGACATCTCTCACTTAATCTTCGGTGTACCTTCACATA 598
D 383 ACTCAAACTCCGAGCTGGCTTTCCAGTGGCCACACAGCATCTATGGAATCCCATCTGTCA 442
QY 599 TAAACACGCAAAATATATGTTTTCAGAGCCATGCAATCTGTATCGCAGCTAACACAA 658
D 443 TCAATCTGCCAATACGTGTATTTCTTGGCTTGGAGAACTCT----- 487
QY 659 AAGAGCCTTTGTATCAATATTTGATTAAGATTTTCAACGAGAAATTTGATCAATCTACATA 718
D 488 TAAACCTTGTATCACCAGATGCGAGTGAAGCTTTTACCGCGCAGCTTTTGGAACTCCATC 547
QY 719 GGGGCAAGGCTTGGATATATCTGAGAGAGCTTTCTGCTGAAATCATACCTACTCAGG 778
D 548 AGGGCAAGGCTAGATATTTACTGGAGGA-----TAATTACCTTGTCCCACTGAAG 602
QY 779 AGATCTATTGATATGTTATGATTAATAAACAAGGCGGCTTTTTCAGATTAAAGTTGAGAC 838
D 603 AAGATATATAAGCTATGTTGCTGCGAAGAAACAGTGGAGCTGTTTGGATTGAGTAGTGT 662
QY 839 TCATGGAAGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 866
D 663 TCATGAGTGTGTTCTCTGATTACCAAGA 690
```

```
RESULT 15
BG675075 974 bp mRNA linear EST 01-MAY-2001
LOCUS 602621320F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746676 5',
DEFINITION mRNA sequence.
ACCESSION BG675075
VERSION BG675075.1 GI:13906471
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 974)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10595 row: h column: 05
High quality sequence stop: 849.
FEATURES
Location/Qualifiers
1..974
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4746676"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5Kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
```

```
Query Match 6.1%; Score 96.2; DB 4; Length 974;
Best Local Similarity 51.4%; Pred. No. 3.3e-12;
Matches 320; Conservative 0; Mismatches 273; Indels 30; Gaps 3;
QY 404 GCAAGAACTTTAGACTAAATTTAATAGTTTCAAAATTAACAGAGTTATGAAATTTGCCAAAG 463
D 45 GTAAACAAGTGAAGAACCAAACTTTCCACAGGCAATTTAATCAATGGCTGAAGTTCCAGAGG 104
QY 464 ACCAGCTGCCCATAGTTTTCGCAAAATTTGAGCTCTTGCATAAATCCAGCCCTTTTAAATCG 523
D 105 ACAAGCTACAGATTAATTTGAAGTGCAGAAATGTTGCAATATGCCAGTTTACTCATCG 164
QY 524 ACGATATAGAAGATAATGCTCCCTTTGAGAAGGGGACAGACCACTTCTCACTTAAATCTTCG 583
D 165 ATGATATTGAGACAACTCAAACTCCGAGCTGGCTTTCCAGTGGCCCAACAGCATCTATG 224
QY 584 GGTGACCTTCCACTATAAACACCCGCAAAATATATGTTTTCAGAGCCATGCACTTTGTAT 643
D 225 GAATCCCATCTGTCATCAATTTCTGCCAATTTAGCTGTTATTTCTGGCTTGGAGAAAGTCT 284
QY 644 CGCAGCTAACCAAAAGAGAGCTTTTGTATCATAAATTTGATTACGATTTTCAACGAGAAAT 703
D 285 TAACTCTTGATCACCAGATGCGAGTGAAGC-----TTTTTACCGCCAGC 329
QY 704 TGATCAATCTCATAGGGGACAAAGCTTTGGATATATATCTGGAGAGACTTTTCTGCTTGAAA 763
D 330 TTTTGGAACTCCATCAGGGACAAAGGCTTAGATATTTACTGGAGGA-----TAATTACA 383
QY 764 TCATACCTACTCAGGAGATGTTTGAATATGTTTATGAATATGTTTGAATATAAAGAGGCGGCTTTTCA 823
D 384 CTTGTCCCACTGAAGAAGAAATATAAAGCTATGGTGTCTGCAGAAACAGGTGGAGCTTTTG 443
QY 824 GATTAACGTTGAGACTCATGGAAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 883
D 444 GATTAGCAGTAGTCTCATGCGAGTTGTTCTCTGATTACAAAGAGATTTTAAACCGCTAC 503
QY 884 TTTCTTTTCAATAATCTTCTGGGTATTTATTTATCATGATTAGAGATGATTACTTGAATTTGA 943
D 504 TT-----AATACACTTGGCTCTTTTCCAAATTAGGATGATTATGCTTAATCTAC 554
QY 944 AAGATTTCCAAATGTCACGCGAAAGAGCTTTTCTGAGGACATTTACAGAGGGGAAAGTTAT 1003
D 555 ACTCCAAAGATATATAGTGAAGGCTTTTGTGAGAGATCTGCAGAGGGGAAAGTTCT 614
QY 1004 CTTTTCCTCATGTCGTCACGCGCTT 1026
D 615 CATTTCTCTACTATTCATGCTATT 637
```

Search completed: October 17, 2004, 03:24:28
Job time : 4884.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:43:59 ; Search time 128.441 Seconds
(without alignments)
8682.774 Million cell updates/sec

Title: US-10-041-018-1
Perfect score: 1569
Sequence: 1 aatattacatagatagatag.....aatgcgcggtaaatcgga 1569
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| 1 | 1569 | 100.0 | 1569 | 2 | US-08-761-344-3 |
| 2 | 1005 | 64.1 | 1005 | 2 | US-08-761-344-1 |
| 3 | 114 | 7.3 | 903 | 1 | US-08-469-665-1 |
| 4 | 114 | 7.3 | 903 | 2 | US-09-038-596-1 |
| 5 | 114 | 7.3 | 903 | 5 | PCT-US95-00421-1 |
| 6 | 70.6 | 4.5 | 1882 | 3 | US-09-091-725-14 |
| 7 | 65 | 4.1 | 498 | 4 | US-09-248-796A-3882 |
| 8 | 52.6 | 3.4 | 832 | 4 | US-09-621-976-2813 |
| 9 | 49.8 | 3.2 | 1141 | 4 | US-09-806-708B-22 |
| 10 | 48.2 | 3.1 | 1664976 | 4 | US-08-916-421B-1 |
| 11 | 48.2 | 3.1 | 1664976 | 4 | US-09-692-570-1 |
| 12 | 45.2 | 2.9 | 392000 | 4 | US-10-027-983-11 |
| 13 | 44.8 | 2.9 | 486 | 4 | US-09-248-796A-3883 |
| 14 | 44.4 | 2.8 | 6583 | 4 | US-10-204-708-25 |
| 15 | 43 | 2.7 | 1055 | 4 | US-09-806-708B-23 |
| 16 | 42.8 | 2.7 | 1141 | 4 | US-09-806-708B-22 |
| 17 | 41.2 | 2.6 | 5152 | 4 | US-10-204-708-73 |
| 18 | 41 | 2.6 | 765 | 4 | US-09-601-198-160 |
| 19 | 40.8 | 2.6 | 1956 | 3 | US-08-559-996B-1 |
| 20 | 40.8 | 2.6 | 1956 | 4 | US-09-351-794A-1 |
| 21 | 40.6 | 2.6 | 6656 | 4 | US-10-204-708-75 |
| 22 | 40 | 2.5 | 993 | 1 | US-08-705-377-2 |
| 23 | 40 | 2.5 | 993 | 1 | US-08-705-377-4 |
| 24 | 40 | 2.5 | 993 | 1 | US-09-052-962-2 |
| 25 | 40 | 2.5 | 993 | 2 | US-09-052-962-4 |
| 26 | 40 | 2.5 | 993 | 2 | US-09-053-068-2 |
| 27 | 40 | 2.5 | 993 | 2 | US-09-053-068-4 |

28 39.8 2.5 687 4 US-09-248-796A-8427 Sequence 8427, Ap
29 39 2.5 832 4 US-09-621-976-2813 Sequence 2813, Ap
30 39 2.5 6020 4 US-10-204-708-7 Sequence 7, Appli
31 39 2.5 17327 1 US-07-906-871-15 Sequence 15, Appl
32 38.8 2.5 3701 3 US-08-845-258-10 Sequence 10, Appl
33 38.8 2.5 3701 3 US-08-990-571-10 Sequence 10, Appl
34 38.8 2.5 3701 3 US-08-723-142A-10 Sequence 10, Appl
35 38.8 2.5 3701 4 US-09-528-784A-10 Sequence 10, Appl
36 38.8 2.5 3701 4 US-09-569-098A-10 Sequence 10, Appl
37 38.6 2.5 4770 3 US-09-000-094-45 Sequence 45, Appl
38 38.6 2.5 4770 4 US-10-011-749-45 Sequence 45, Appl
39 38.6 2.5 4770 4 US-09-000-004-45 Sequence 45, Appl
40 38.4 2.4 990 1 US-08-410-167A-1 Sequence 1, Appli
41 38.4 2.4 993 1 US-08-705-377-1 Sequence 1, Appli
42 38.4 2.4 993 1 US-08-705-377-3 Sequence 3, Appli
43 38.4 2.4 993 1 US-08-705-377-6 Sequence 6, Appli
44 38.4 2.4 993 2 US-09-052-962-1 Sequence 1, Appli
45 38.4 2.4 993 2 US-09-052-962-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-761-344-3
; Sequence 3, Application US/08761344
; Patent No. 5912154
; GENERAL INFORMATION:
; APPLICANT: Ferro-No. 5912154ick, Susan
; APPLICANT: Jiang, Yu
; TITLE OF INVENTION: GERANYLGERANYL DIPHOSPHATE SYNTHASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/761,344
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 30,020
; REFERENCE/DOCKET NUMBER: 3161-14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/862-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 1569 base pairs
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-761-344-3

Query Match 100.0%; Score 1569; DB 2; Length 1569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATATTACATAGATAGGACAGCCGCGGATTTTTCATCTCAAGAGGTAATCTTATT 60
Db 1 AATATTACATAGATAGGACAGCCGCGGATTTTTCATCTCAAGAGGTAATCTTATT 60

NAME/KEY: CDS
LOCATION: 1..1005
US-08-761-344-1

Query Match
Best Local Similarity 64.1%; Score 1005; DB 2; Length 1005;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ATGGAGGCCAAGATGATGAGCTGATCAATAATGATCCCTGTTGGTCCAGCCAAATGAA 360
Db
1 ATGGAGGCCAAGATGATGAGCTGATCAATAATGATCCCTGTTGGTCCAGCCAAATGAA 60
QY 361 ACCTTGATTTCAAACCTTATATACATCCCTTTGAAACCTGGCAAGAACTTTAGACTA 420
Db
61 ACCTTGATTTCAAACCTTATATACATCCCTTTGAAACCTGGCAAGAACTTTAGACTA 120
QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 480
Db
121 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 180
QY 481 TCGCAAAATGTTGAGCTCTTGCATATATCCAGCCCTTTTAATCGACGATATAGAATAAT 540
Db
181 TCGCAAAATGTTGAGCTCTTGCATATATCCAGCCCTTTTAATCGACGATATAGAATAAT 240
QY 541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
Db
241 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 300
QY 601 AACACCGCAAAATATATGATTTTACAGCCATGCACTGTTGATCGCAGCTAACCAAAA 660
Db
301 AACACCGCAAAATATATGATTTTACAGCCATGCACTGTTGATCGCAGCTAACCAAAA 360
QY 661 GAGCCTTTGATCATFAATTTGATTTACGATTTTCAACGAAGAAATGATCAATCTCATATAGG 720
Db
361 GAGCCTTTGATCATFAATTTGATTTACGATTTTCAACGAAGAAATGATCAATCTCATATAGG 420
QY 721 GGCAAGCCTTGGATATATATCTGAGAGACTTCTGCTGAAATCATACCTACTCAGGAG 780
Db
421 GGCAAGCCTTGGATATATATCTGAGAGACTTCTGCTGAAATCATACCTACTCAGGAG 480
QY 781 ATGATTTGATATGTTGATGATTAATAAAGAGGCTTTTCAAGTAACTTCAAGAAATTT 840
Db
481 ATGATTTGATATGTTGATGATTAATAAAGAGGCTTTTCAAGTAACTTCAAGAAATTT 540
QY 841 ATGGAAGCGCTGTCTCTCTCTCTCACACCGCCATTCGTTGGTTCCTTTCATAAATCTT 900
Db
541 ATGGAAGCGCTGTCTCTCTCTCTCACACCGCCATTCGTTGGTTCCTTTCATAAATCTT 600
QY 901 CTGGGTATTTATCATGATTTAGAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTC 960
Db
601 CTGGGTATTTATCATGATTTAGAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTC 660
QY 961 AGCGAAAAGGCTTTGCTGAGGACATATACAGAGGGAGTATCTTTCCCATCGTCCAC 1020
Db
661 AGCGAAAAGGCTTTGCTGAGGACATATACAGAGGGAGTATCTTTCCCATCGTCCAC 720
QY 1021 GCCTTTAACTTCAATAAAGGATGATTTCAAACTGAGCAACATGAAATTTCTAAGAAAT 1080
Db
721 GCCTTTAACTTCAATAAAGGATGATTTCAAACTGAGCAACATGAAATTTCTAAGAAAT 780
QY 1081 CTCTGTTGAGGACAACTGATTAAGATATAAATAAAGTGAATTTCAAACTGGAATTC 1140
Db
781 CTCTGTTGAGGACAACTGATTAAGATATAAATAAAGTGAATTTCAAACTGGAATTC 840
QY 1141 GACACCAATTCATTTGGCTCACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200
Db
841 GACACCAATTCATTTGGCTCACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 900
QY 1201 AATGATAATGAAAATAAGTATTTACCTGATTTGGCTTCGATTTCCGATCCGACCGCCACCAAT 1260
Db
901 AATGATAATGAAAATAAGTATTTACCTGATTTGGCTTCGATTTCCGATCCGACCGCCACCAAT 960
QY 1261 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAATTTG 1305

Db 961 TTACATGACGAATTTGTTATATATATATAGACCACTTATCCGAATTTG 1005

RESULT 3

US-08-469-665-1
Sequence 1, Application US/08469665
Patent No. 5786193

GENERAL INFORMATION:

APPLICANT: GREENE, ET AL.
TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA: US/08469,665
FILING DATE: June 6, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00421
FILING DATE: 11 JAN 95
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-377
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-469-665-1

Query Match 7.3%; Score 114; DB 1; Length 903;
Best Local Similarity 52.1%; Pred. No. 3e-20;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAAGCTTGATTTCAAACCTTATAATCACATCCCTTTTGAACCTTGGCAAGAACTTTAGAC 418
Db 26 AAAGAATTTCTTAGAACCTTATAATTAATCTTCTTCACTTACCAAGTAAACAGTGAAGAA 85
QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTAATTTGCCAAAGACCAAGCTGGCCATAG 478
Db 86 CCAAACTTTTCCAGGCAATTTAATCAITGGCTGAAAGTTCCAGAGGACAAAGCTACAGATTA 145
QY 479 TTTTCGAAAATTTGTTGAGCTCTTCGCAATTTCCAGCCCTTTTAAATCGACGATATAGAAGATA 538
Db 146 TTATTGAGTGACAGAAAATGTTGCAATTAATGCCAGTTTACTCATCTGATGATTAATGAAGACA 205
QY 539 ATGCTCCCTTTGAGAAGGGGACAGACCACTTCTCACTTTAAATCTTCGGTGTACCTCCACTA 598
Db 206 ACTCAAACTTCCCGACGTTGGCTTTCCAGTGGCCCAAGCATCTATGATATCCATCTCTGCA 265
QY 599 TAAACACCGCAATTTATATGTTTATTTTTCAGAGCCCAAGCAACTTTGTATCGCAGCTAACCAAA 658
Db 266 TCAATTTCTGCAATTAACGTTGATTTTCTTGGCTTGGAGAAAGTCT----- 310
QY 659 AAGAGCCTTTGTTATCATATTTGATTTACGATTTTCAACGAGAAATTTGATCAATCTACATA 718

```

Db      311 TAACCTTTGATCACCACGATGAGTGAAGCTTTTACCGCAGCTTTTGGAACTCCATC 370
Qy      719 GGGGACAGGCTTGTGATATATACCTGGAGAGCTTTTCTGCTGAAATCATACCTACTCAGG 778
Db      371 AGGGACAGGCTTGTGATATATACCTGGAGAGCTTTTCTGCTGAAATCATACCTACTCAGG 424
Qy      779 AGATGATTTTGAATGTTTATGAAATATAAAGGGGGCTTTTTCAGATTAAGCTTGGAGAC 838
Db      425 AAGATATAAAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 484
Qy      839 TCATGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
Db      485 TCATGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 535
Qy      899 TTCTGGGTATTTATCAGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 958
Db      536 CACTTGGGCTCTTTTCCAAATAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 595
Qy      959 CCAGCGAAAAAGGCTTGTGCTGAGGACATTTACAGAGGGGAGTATTTTCCCATCGTCC 1018
Db      596 GTGAAAAACAAAAGTTGGGTGAGATCTGACAGAGGGGAAAGTTCTCATTTCTCTATTTC 655
Qy      1019 ACGCCCTT 1026
Db      656 ATGCTATT 663

```

```

RESULT 4
US-09-038-596-1
; Sequence 1, Application US/09038596
; Patent No. 5928924
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA: US/09/038,596
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: cDNA
; US-09-038-596-1

```

```

Query Match      7.3%; Score 114; DB 2; Length 903;
Best Local Similarity 52.1%; Pred. No. 3e-20;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
Qy      359 AAAGCTTGATTTCAAACCTTATATATCATCTCTTTTGAACCTTGGCAAGTATTAGAC 418
Db      26 AAAGAAATCTTCTAGAACCTTATAATATCTTACTTACGTTACCAAGTAAACAAGTGAGAA 85
Qy      419 TAAATTTTAAAGTTCAAATTAACAGATTTATGAATTTGCCAAAGACACAGCTGGCCATAG 478
Db      86 CCAAATTTTCAAGGCAATTTAATCATGTGCTGAAAGTTCAGAGACAGCTACAGATTA 145
Qy      479 TTTTCGCAATTTGCTGAGCTCTTTCGCAATTTTCCAGCGCTTTTAAATCGACGATATAGAATA 538
Db      146 TTAATGAGTGCACAGAAATGTTGCATAATGCCAGTTTACTCATGATGATATGAAGACA 205
Qy      539 ATGCTCCCTTGAGAGAGGGGACAGACCATTTCTCATTTAATCTTCGTTGTTACCTTCCACTA 598
Db      206 ACTCAAAACTCCGAGCTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 265
Qy      599 TAAACACCGCAATTTATGATTTTCAGAGCCATGCACTTGTATCGCAGCTAAACACCAA 658
Db      266 TCAATTTCTGCCAATTTACGTTGATTTTCTTGGCTTGGAGAAAGTCT----- 310
Qy      659 AAGAGCCTTTGTATCATTAATTTGATTACGATTTTCAACGAGAAATTTGATCAATCTACATA 718
Db      311 TAAACCTTTGATCACCAGATGAGTGAAGCTTTTACCAGGCTTTTGGAACTCCATC 370
Qy      719 GGGGACAGGCTTGGATATATATCTGGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778
Db      371 AGGGACAGGCTTAGATATTTTCTGGAGGGA-----TAAATTACACTTGTCCCACTGAAG 424
Qy      779 AGATGATTTTGAATGTTTATGATTAATAAACAAGGCGGCTTTTTCAGATTAACGCTTGAGAC 838
Db      425 AAGATATAAAGCTATGTTGCTGAGAGAAACAGGTGACTGTTTGGATTAGCAGTAGGTGTC 484
Qy      839 TCATGGAAGCGCTGCTCTCTCTCTCACACACGCGCAATTCGTTGGTTCCTTTTCATTAATC 898
Db      485 TCATGCGATTTGTTCTCTGATTTACAAAGAAATTTTAAACCGCTACTT-----ATA 535
Qy      899 TTCTGGGTATTTATCATGATTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 958
Db      536 CACTTGGGCTCTTTTCCAAATAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 595
Qy      959 CCAGCGAAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAGTATTTTCCCATCGTCC 1018
Db      596 GTGAAAAACAAAAGTTGGGTGAGATCTGACAGAGGGGAAAGTTCTCATTTCTCTATTTC 655
Qy      1019 ACGCCCTT 1026
Db      656 ATGCTATT 663

```

```

RESULT 5
PCT-US95-00421-1
; Sequence 1, Application PC/TUS9500421
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS

```


QY 592 TCCACTATAACACCGCAAAATATATGTTTTCAGAGCCATGCAACTTGTATCGCAGCTA 651
 |||||
 Db 370 CAGACAATAAACACTGCAAACTACGCTACTTCTGTGGCTTATCAAGATCTTCAAGCTT 429
 |||||
 QY 652 ACCACAAA 660
 |||||
 Db 430 CGCCCAACA 438

RESULT 7
 US-09-248-796A-3882
 ; Sequence 3882, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 3882
 ; LENGTH: 498
 ; TYPE: DNA
 ; ORGANISM: Candida albicans
 ; ORGANISM: Candida albicans
 US-09-248-796A-3882

Query Match 4.1%; Score 65; DB 4; Length 498;
 Best Local Similarity 58.8%; Pred. No. 1.5e-07;
 Matches 151; Conservative 0; Mismatches 100; Indels 6; Gaps 2;
 QY 784 TATTGGAATGTTTATGAATAAACAAGCGGCTTTTCAGATTAACTGAGACTCATG 843
 |||||
 Db 10 TATTGGAATGTTTATGAATAAACAAGCGGCTTTTCAGATTAACTGAGACTCATG 69
 |||||
 QY 844 GAAGCGCTGCTCTCTTCTCACACCAGCGGCATTCGTTGCTTCTTCAATAATCTCG 903
 |||||
 Db 70 CTGCTGTAATCAGATGTCACAGAAACG--ATCAATTAATATCCCTTGCAAAATTTAATG 126
 |||||
 QY 904 GGTATATTATCAGATTAGATGATCTTCAATTTGAAGATTTCAAAATGTCAGC 963
 |||||
 Db 127 GGCATTTTGTATCAAGTTAGAGACGATTAATTCGAATTTAGTGGATGCCAGTATCTGCG 186
 |||||
 QY 964 GAAAGAGGCTTTGC--TGAGGACATTACAGAGGGAAGTTATCTTTTCCATCGTCCAC 1020
 |||||
 Db 187 ATGAAGGCACTACATGTTGAAGCTTGAAGGAGGAAAGTTATCAATTCCTATATACAT 246
 |||||
 QY 1021 GCCCTTAATCTCACTAA 1037
 |||||
 Db 247 TGTCTAAGAACCACTAA 263

RESULT 8
 US-09-621-976-2813/c
 ; Sequence 2813, Application US/09621976
 ; Patent No. 6639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 2813
 ; LENGTH: 832

; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 235..399
 ; US-09-621-976-2813
 Query Match 3.4%; Score 52.6; DB 4; Length 832;
 Best Local Similarity 16.0%; Pred. No. 0.00033;
 Matches 45; Conservative 133; Mismatches 102; Indels 1; Gaps 1;
 QY 1268 AGCAATTTTATATATAATAGACCACTTATCCGAATTTGAAATAAATGATCAATCAAA 1327
 |||||
 Db 393 AAGAATTTGACCAATAAATAATATATTTTGTGTTTWWKTYWYTTTTRMWWMMKKARR 334
 |||||
 QY 1328 TTAGTGGAGGAGATAGTCAGAAATAA-AGCCTTCTCTCTCTCTCTCTCTCTCTCTCT 1386
 |||||
 Db 333 WYWKWSTYACASRYRKYTWGWWWYWKRMSTRYCYMCWCKCMYGRRCAYWYMARGR 274
 |||||
 QY 1387 TACGATTTTATATATATACGTTTCATTTGCATCATCTTTTGTATATATCTCAAAAAGATCTTT 1446
 |||||
 Db 273 WMSYAWGKWKMSRMSMSMCTRMYYKKGSTYWTMKTCTATWCYWKYKRMWMSKTCWSGS 214
 |||||
 QY 1447 AGTTCGCAAAATAGTCAAAATCTTCAAAATTTATAGCCTTTATATATTTTCCACGATTTCTGA 1506
 |||||
 Db 213 RGGYMTSYTSTRSYSNYASWNYTACWGWGRWSTYTWYMAWGWKMYRYATTWRRAMWWW 154
 |||||
 QY 1507 AACTCCTTTTATACGCGCTTAAATGCTAGCGGTTACTGT 1547
 |||||
 Db 153 AAATWYWWYMAWCMSSRGAAMRYRRIMMWGYRYWWRKASY 113

RESULT 9
 US-09-806-708B-22
 ; Sequence 22, Application US/09806708B
 ; Patent No. 6784342
 ; GENERAL INFORMATION:
 ; APPLICANT: The University of British Columbia
 ; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
 ; FILE REFERENCE: 4810-58741
 ; CURRENT APPLICATION NUMBER: US/09/806,708B
 ; CURRENT FILING DATE: 2001-04-03
 ; PRIOR APPLICATION NUMBER: US 60/147,133
 ; PRIOR FILING DATE: 1999-08-04
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 1141
 ; TYPE: DNA
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; NAME/KEY: promoter
 ; LOCATION: (1)..(1141)
 ; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEL promoters
 ; US-09-806-708B-22

Query Match 3.2%; Score 49.8; DB 4; Length 1141;
 Best Local Similarity 12.7%; Pred. No. 0.002;
 Matches 97; Conservative 257; Mismatches 408; Indels 2; Gaps 2;
 QY 295 AAATCAATGAGGCCAAGATAGATGAGCTGATCAATAATGATCTGTTGGTCCAGCAA 354
 |||||
 Db 377 WHANAAYSRKKWTBYKRKTNNNNNGTTWKKEMWAMYFDDWDBGTNNNNNGGRTYY 436
 |||||
 QY 355 AATGAAAGCTTGATTTCAAAAACCTTATAATACATCACTCTTTTGAACCTGCGCAAGACTTT 414
 |||||
 Db 437 GWTKNKKOYTYKWKANNCKWRANDHKTCTHNNYTWKMTYNNVNCVYKSWTKSHRBA 496
 |||||
 QY 415 AGACTAAATTAATAGTTCAAAATTAACAGAGTTATGAAATTTGGCCAAAGACCGAGTCC 474
 |||||
 Db 497 AAVVTWYWWRRYAHANNNDWYWKACTYKYBVCWKWNNYAAWYTKSSMNTSRY 556
 |||||
 QY 475 ATAGTTCCGCAAAATGTTGAGCTTTCGATAAATCCAGGCTTTTAAATCGACGATATAGAA 534

[illegible]


```

; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103398)..(103398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)

```

```

; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g

```

```
FEATURE: misc feature
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g

Query Match      3.1%; Score 48.2; DB 4; Length 1664976;
Best Local Similarity 48.1%; Pred. No. 0.11;
Matches 137; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 368 TTTCAAAACCTTATATACATCATCTTTTGAACCTGGCGAAGACATTAGACATAATTTAA 427
Db 688622 TATATAACCGTCAAAACATCTTCTATTGCTGGAGGAAGAGAAATAGGCCATATTTAA 688681
QY 428 TAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTTTCGCAA 487
Db 688682 CTGTAGTAACTTATATGTTGGAAGAACGATATTTGAGGAGGTTTTCGCGCGCTGCTG 688741
QY 488 TTGTTGAGCTCTTGCAATAATCCAGCCTTTTAATGACGATATAGAGATAATGCTCCCT 547
Db 688742 CAGTAGAGTTAATTCACAACTACACCTTAATACATGATGACATATGGAATGATGATG 688801
QY 548 TGAGAGGGGACAGACCACTTCTACCTTAATCTCGTGTAACCTCCACTATATAACACCG 607
Db 688802 AGAGGAGAGGAAACCAACAGTTTCATGTTGCTATGGAGAGCCAAATGGCTATCTTAGCTG 688861
QY 608 CAAATTATATGTTATTTTCAGAGCCATGCAACTTGTATCGCAGCTAA 652
Db 688862 GAGATTATATGTTAAGCTTTTGAGCAGTTTCAAGATAA 688906
```

RESULT 12

```
US-10-027-983-11
Sequence 11, Application US/10027983
Patent No. 6617162
GENERAL INFORMATION:
APPLICANT: Kenneth W. Dobie
APPLICANT: Mark P. Roach
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
FILE REFERENCE: RTS-0340
CURRENT APPLICATION NUMBER: US/10/027,983
CURRENT FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 11
LENGTH: 392000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 137740
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 137742
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (138122)..(138221)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 145507
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 151967
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (151967)..(1542066)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 154217
OTHER INFORMATION: unknown
NAME/KEY: misc feature

LOCATION: (164037)....(164136)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (174657)....(174756)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (186224)....(186323)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (195242)....(195341)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 202703
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (202771)....(202870)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (206246)....(215602)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (218126)....(218225)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (220360)....(220459)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (222717)....(222816)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (223981)....(224080)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (227487)....(227586)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (230157)....(230256)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (232299)....(232398)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (236552)....(236651)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (238789)....(248788)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: exon
LOCATION: (118288)....(119101)
OTHER INFORMATION: exon 1C
NAME/KEY: exon:intron junction
LOCATION: (151129)....(151130)
OTHER INFORMATION: exon 5:intron 5
NAME/KEY: exon:intron junction
LOCATION: (299248)....(299249)
OTHER INFORMATION: exon 9:intron 9
NAME/KEY: exon:intron junction
LOCATION: (348578)....(348579)
OTHER INFORMATION: exon 10:intron 10
NAME/KEY: intron
LOCATION: (348579)....(381838)
OTHER INFORMATION: intron 10
NAME/KEY: intron:exon junction
LOCATION: (386185)....(386186)
OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

Query Match      2.9%; Score 45.2; DB 4; Length 392000;
Best Local Similarity 50.0%; Pred. No. 0.36;
Matches 113; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 598 ATAACACCGCAATATATGTTATTCAGAGCCATGCACTTGTATCGCAGCTAACACACA 657
```

```

RESULT 14
US-10-204-708-25/c
; Sequence 25, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
;

```

| | Query Match | 2.7% | Score 43; | DB 4; | Length 1055; |
|----|-----------------------------------|---|-----------------|-----------------|--------------------|
| | Best Local Similarity | 21.3%; | pred. No. 0.12; | Mismatches 157; | Indels 12; Gaps 2; |
| | Matches 126; | Conservative 157; | Mismatches 299; | Indels 12; | Gaps 2; |
| Qy | 886 | CCTTTCAATAAATCTTCGGTATTATTATCAGATTAGAGATGATTACTTGAAATTTGAAA | 945 | | |
| | : : : : : | | | | |
| Dd | 634 | CATNTWARATNNTTRAGCAAINTKGTSTTAGMWTGATTGNNNNNNNNNCMTASC | 575 | | |
| | : : : : : | | | | |
| Qy | 946 | GATTTCCAANTGCCACGCGAAAAAGCTTTCTGAGGACAATTACAGAGGGGAAGTTATCT | 1005 | | |
| | : : : : : | | | | |
| Dd | 574 | CTCTGACTAATTATRGATSACKRKTAWTTCTYRRRWANGWRWTAANMFCAWRTYTRY | 515 | | |
| | : : : : : | | | | |

us-10-041-018-1.rni

Search completed: October 17, 2004, 03:30:33
Job time : 132.441 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 16:35:39 ; Search time 733.231 Seconds
(without alignments)
11232.937 Million cell updates/sec

Title: US-10-041-018-1

Perfect score: 1569

Sequence: 1 aatattacatatagatag.....aatcgccggttaattcgga 1569

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_23Sep04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002as:*

7: Geneseq2002bs:*

8: Geneseq2003as:*

9: Geneseq2003bs:*

10: Geneseq2003cs:*

11: Geneseq2003ds:*

12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----|--------------------|
| 1 | 1569 | 100.0 | 1569 | 12 | Adm98581 Geranylge |
| 2 | 1008 | 64.2 | 1008 | 6 | AAL40786 DNA relat |
| 3 | 1008 | 64.2 | 1008 | 6 | Abk96782 S. cerevi |
| 4 | 1008 | 64.2 | 1008 | 10 | Adc49385 Yeast BTS |
| 5 | 1008 | 64.2 | 1008 | 12 | Adl14810 Yeast ger |
| 6 | 568.6 | 36.2 | 599 | 10 | Adk52980 Plant DNA |
| 7 | 183.4 | 11.7 | 963 | 10 | Adf75080 A. gossyp |
| 8 | 114 | 7.3 | 903 | 2 | Aat31846 Human ger |
| 9 | 114 | 7.3 | 903 | 2 | Aav41555 Human ger |
| 10 | 114 | 7.3 | 903 | 2 | Aax86773 Human ger |
| 11 | 114 | 7.3 | 1395 | 12 | Adj10255 Human ger |
| 12 | 114 | 7.3 | 1547 | 3 | Aaz60889 cDNA enco |
| 13 | 114 | 7.3 | 1692 | 5 | Aaz60890 cDNA enco |
| 14 | 114 | 7.3 | 2534 | 12 | Adl63550 Human ova |
| 15 | 114 | 7.3 | 2534 | 12 | Adm98583 Geranylge |
| 16 | 113.6 | 7.2 | 2529 | 3 | Aaf18152 Lung canc |
| 17 | 99.6 | 6.3 | 1032 | 8 | Abz77169 Phoma bet |
| 18 | 99.6 | 6.3 | 1888 | 5 | Abv25136 Human pro |
| 19 | 87.6 | 5.6 | 17131 | 3 | Aaz60888 DNA enco |
| 20 | 87.6 | 5.6 | 51001 | 12 | Adj10262 Human ger |
| 21 | 84.2 | 5.4 | 2210 | 4 | ABL08261 Drosophil |

| | | | | | | |
|------|------|-----|--------|----|-------------|--------------------|
| 22 | 84.2 | 5.4 | 2222 | 12 | Adm98601 | Geranylge |
| 23 | 74.8 | 4.8 | 630 | 10 | ADD34746 | Mouse mit |
| 24 | 70.6 | 4.5 | 1882 | 2 | AAT72942 | Phaffia c |
| c 25 | 68.8 | 4.4 | 554 | 6 | ABQ57629 | Human col |
| c 26 | 68.4 | 4.4 | 716 | 6 | ABQ58229 | Human col |
| c 27 | 65.6 | 4.2 | 593 | 6 | ABQ57793 | Human col |
| 28 | 62.4 | 4.0 | 1843 | 12 | Adm98585 | Geranylge |
| 29 | 60 | 3.8 | 399 | 5 | ABV49901 | Human pro |
| 30 | 58.6 | 3.7 | 1095 | 8 | ACA52518 | Prokaryot |
| c 31 | 56.4 | 3.6 | 6115 | 6 | ABL33801 | Human imm |
| c 32 | 49.2 | 3.1 | 6802 | 6 | ABL33231 | Human imm |
| c 33 | 48.8 | 3.1 | 1044 | 8 | ACA23035 | Prokaryot |
| c 34 | 48.8 | 3.1 | 110000 | 2 | RAX20248_05 | Continuation (6 of |
| c 35 | 48.8 | 3.1 | 111309 | 2 | RAX20250_06 | Continuation (7 of |
| 36 | 48.2 | 3.1 | 110000 | 2 | AAV21209_06 | Continuation (7 of |
| 37 | 48 | 3.1 | 2000 | 8 | ADA71938 | Rice gene |
| c 38 | 47.8 | 3.1 | 6127 | 6 | ABL33614 | Human imm |
| c 39 | 47.8 | 3.0 | 2703 | 6 | ABN67916 | Human imm |
| c 40 | 47.8 | 3.0 | 110000 | 6 | ABN71527_13 | Streptoco |
| c 41 | 47 | 3.0 | 383 | 5 | ADL44116 | Continuation (14 o |
| c 42 | 46.4 | 3.0 | 6681 | 6 | ABL32155 | Human ova |
| c 43 | 46.4 | 3.0 | 6681 | 6 | ABL54304 | Human imm |
| c 44 | 46.4 | 3.0 | 10856 | 6 | ABL32457 | Chemical |
| c 45 | 46.4 | 3.0 | 18133 | 6 | ABK40018 | Human che |

ALIGNMENTS

RESULT 1

Adm98581

ID Adm98581 standard; DNA, 1569 BP.

XX Adm98581;

AC Adm98581;

DT 01-JUL-2004 (first entry)

XX Geranylgeranyl pyrophosphate synthase DNA #1.

DE Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;

XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;

KW diterpene precursor; diterpene synthase; defence toxin;

KW volatile defensive signal; pollinator attractant; photoprotectant.

XX Saccharomyces cerevisiae.

OS US2004072323-A1.

XX 15-APR-2004.

PD 07-JAN-2002; 2002US-00041018.

PF 05-JAN-2001; 2001US-0259880P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

PI Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

DR New unicellular organisms comprising exogenous nucleic acids encoding a

XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for

XX producing diterpenes and diterpene precursors.

PS Claim 2; SEQ ID NO 1; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene

XX or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a

CC promoter operable in the organism, and an exogenous nucleic acid sequence

CC encoding a diterpene synthase under the control of a promoter operable in

CC the organism. The invention also relates to methods of producing a

CC diterpene or diterpene precursor and a method of isolating a diterpene

CC synthase comprising growing several cells in the presence of a
 CC polyaromatic resin to make a cell/resin mixture, where at least one of
 CC the cells further comprises at least one isolated and purified nucleic
 CC acid sequence of a yeast expression library, and the expression of the
 CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents DNA encoding a geranylgeranyl pyrophosphate synthase
 CC polypeptide of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 1569 BP; 516 A; 307 C; 260 G; 486 T; 0 U; 0 Other;

Query Match 100.0%; Score 1569; DB 12; Length 1569;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATATTACATATAGATATAGACCAAGCCCGCATTTTTCATCTCAAGAGTAAATCTTATT 60
 DB 1 AATATTACATATAGATATAGACCAAGCCCGCATTTTTCATCTCAAGAGTAAATCTTATT 60

QY 61 ATTATAGTGGTATCCAAAGCTTCAACGCTTCCAGCATAGACAGAAATTCATGTTTGGAT 120
 DB 61 ATTATAGTGGTATCCAAAGCTTCAACGCTTCCAGCATAGACAGAAATTCATGTTTGGAT 120

QY 121 ATGTTATGCTGATCATTTGATCTTACTACCAATTTTCTTCTGCTGCGCTTCTCTTTT 180
 DB 121 ATGTTATGCTGATCATTTGATCTTACTACCAATTTTCTTCTGCTGCGCTTCTCTTTT 180

QY 181 GACGTTTTTTTGAAGCAAAAAAGTCAAGACAGATGCTTTTCAAAAAACCATGTAAAGC 240
 DB 181 GACGTTTTTTTGAAGCAAAAAAGTCAAGACAGATGCTTTTCAAAAAACCATGTAAAGC 240

QY 241 TCATTTTCAAGAGCTACTATATAGAAAGAGAAAGAGTTTACAGTCTGGAAATCA 300
 DB 241 TCATTTTCAAGAGCTACTATATAGAAAGAGAAAGAGTTTACAGTCTGGAAATCA 300

QY 301 ATGAGGCCCAAGATAGATGATCAATATGATCCTGTTTGGTCCAGCCCAAAATGAA 360
 DB 301 ATGAGGCCCAAGATAGATGATCAATATGATCCTGTTTGGTCCAGCCCAAAATGAA 360

QY 361 AGCTTGATTTCAAAACCTTTATATACATCCTTTTGAACCTTGGCAAGAACCTTAGACTA 420
 DB 361 AGCTTGATTTCAAAACCTTTATATACATCCTTTTGAACCTTGGCAAGAACCTTAGACTA 420

QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 480
 DB 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 480

QY 481 TCGCAATTTGTCAGCTCTTGATATATCCAGCTTTTATATCGACGATATAGAGATAAT 540
 DB 481 TCGCAATTTGTCAGCTCTTGATATATCCAGCTTTTATATCGACGATATAGAGATAAT 540

QY 541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACCTCCACTATA 600
 DB 541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACCTCCACTATA 600

QY 601 AACACCGCAATTTATATATTTTACAGGCTTATGATCGAGCTTAAACCAAAAA 660
 DB 601 AACACCGCAATTTATATATTTTACAGGCTTATGATCGAGCTTAAACCAAAAA 660

QY 661 GAGCCTTTGATCATATTTGATTTTACAGGCTTATGATCGAGCTTAAACCAAAAA 720
 DB 661 GAGCCTTTGATCATATTTGATTTTACAGGCTTATGATCGAGCTTAAACCAAAAA 720

QY 721 GGAACAAGGCTTGGATATATATCTGGAGAGACTTCTCGCTGAAATCATCTACTCAGGAG 780

DB 721 GGAACAAGGCTTGGATATATATCTGGAGAGACTTCTCGCTGAAATCATCTACTCAGGAG 780
 QY 781 ATGTTATTTGAATATGCTTATGAATAAACAAGCGGCTTTTTCAGATTAACTGAGACTC 840
 DB 781 ATGTTATTTGAATATGCTTATGAATAAACAAGCGGCTTTTTCAGATTAACTGAGACTC 840
 QY 841 ATGGAAGCGCTGCTCTCTCTTCAACCAAGCGCCATTCGTTGGTTCCTTTTCAATAATCTT 900
 DB 841 ATGGAAGCGCTGCTCTCTCTTCAACCAAGCGCCATTCGTTGGTTCCTTTTCAATAATCTT 900

QY 901 CTGGGTTATTTATCAGATATAGATGATTTACTTGAATTTGAAAGATTTTCCAAATGTC 960
 DB 901 CTGGGTTATTTATCAGATATAGATGATTTACTTGAATTTGAAAGATTTTCCAAATGTC 960

QY 961 AGCGAAAAGGCTTGTCTGAGACATTTACAGAGGGGAGTATCTTTTCCCATCTGTCAC 1020
 DB 961 AGCGAAAAGGCTTGTCTGAGACATTTACAGAGGGGAGTATCTTTTCCCATCTGTCAC 1020

QY 1021 GCCCTTAACTTCACTAAAAACGAAAGTCAAACTGAGCAACACAAATCAAAATTTCAAGAA 1080
 DB 1021 GCCCTTAACTTCACTAAAAACGAAAGTCAAACTGAGCAACACAAATCAAAATTTCAAGAA 1080

QY 1081 CTCTGTTGAGCAAGAGTATTAAGATATATAAAGTCTAAAGCTGATTTCAATCTGGAATTC 1140
 DB 1081 CTCTGTTGAGCAAGAGTATTAAGATATATAAAGTCTAAAGCTGATTTCAATCTGGAATTC 1140

QY 1141 GACACCAATTCATTCGCTTACACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200
 DB 1141 GACACCAATTCATTCGCTTACACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200

QY 1201 AATGATAATGAAATAAGTATTTTACCTGATTTGGCTTCCGACACCGCCCAACCAAT 1260
 DB 1201 AATGATAATGAAATAAGTATTTTACCTGATTTGGCTTCCGACACCGCCCAACCAAT 1260

QY 1261 TTACATGAGCAATTTGTTATATATATATAGACCACTTATCCGAATTTGGAATTAATTTGATC 1320
 DB 1261 TTACATGAGCAATTTGTTATATATATATAGACCACTTATCCGAATTTGGAATTAATTTGATC 1320

QY 1321 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATAAAGCCCTTCTCTCTCTCTTCCGATC 1380
 DB 1321 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATAAAGCCCTTCTCTCTCTTCCGATC 1380

QY 1381 TATACATACGATTTTCAAT 1440
 DB 1381 TATACATACGATTTTCAAT 1440

QY 1441 TCTCTTAGTTCGCAATATAGTCAAAATTTTCAAAATTTTATAGCCCTTATATTTTCCAGAT 1500
 DB 1441 TCTCTTAGTTCGCAATATAGTCAAAATTTTCAAAATTTTATAGCCCTTATATTTTCCAGAT 1500

QY 1501 TTCTGAAACTCTCTTTTATCAGCACCGTTAATGTAGCGGTTACTCTCAAAATCGCCGGTA 1560
 DB 1501 TTCTGAAACTCTCTTTTATCAGCACCGTTAATGTAGCGGTTACTCTCAAAATCGCCGGTA 1560

QY 1561 AATTCGCGA 1569
 DB 1561 AATTCGCGA 1569

RESULT 2

AAL40786

ID AAL40786 standard; DNA; 1008 BP.

XX

AC AAL40786;

XX

DT 03-OCT-2002 (first entry)

XX

DE DNA relating to the production of prenyl alcohol SEQ ID No 5.

XX

KW Prenyl alcohol; mutated cell; squalene synthase gene; geometric isomer;

XX industrial synthesis; isoprenoid-terpenoid compound; gene; ds.

XX

Saccharomyces cerevisiae.

WO200253747-A1.

11-JUL-2002.

20-DEC-2001; 2001WO-JP011215.

28-DEC-2000; 2000JP-00401701.

28-DEC-2000; 2000JP-00403067.

18-SEP-2001; 2001JP-00282978.

(TOYT) TOYOTA JIDOSHA KK.

Ohto C, Obata S;

WPI; 2002-548086/58.

P-PSDB; AAO22195.

Production of prenyl alcohols by culturing translationally-active mutated

cells with reduced squalene synthase gene to express less transcriptional

product.

Disclosure; Page 143-146; 262pp; Japanese.

The invention relates to a method for producing prenyl alcohol comprising

culturing mutated cells, having been mutated so as to reduce the amount

of transcriptional product of squalene synthase gene transcriptional

activity, and then collecting prenyl alcohol from the culture medium. The

method is for the production of prenyl alcohols, which is for use in

industrial synthesis of isoprenoid-terpenoid compounds particularly

physiologically-active prenyl alcohol geometric isomers. This

polynucleotide sequence represents a DNA sequence encoding a protein

relating to the method for producing prenyl alcohol comprising culturing

mutated cells

Query Match 64.2%; Score 1008; DB 6; Length 1008;

Best Local Similarity 100.0%; Pred. No. 9.4e-229;

Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ATGGAGGCCAAGATAGATGAGTCATCAATATGATCCTGTTTGGTCCAGCCAAATGAA 360

Db 1 ATGGAGGCCAAGATAGATGAGTCATCAATATGATCCTGTTTGGTCCAGCCAAATGAA 60

QY 361 AGCTTGATTTCAAACCTTATATACATCCTTTTGAAACCTGGCAAGACTTTAGACTA 420

Db 61 AGCTTGATTTCAAACCTTATATACATCCTTTTGAAACCTGGCAAGACTTTAGACTA 120

QY 421 AATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCAAGACCAGCTGGCCATAGTT 480

Db 121 AATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCAAGACCAGCTGGCCATAGTT 180

QY 481 TCGCAAAATTTGAGCTCTTGCATAATTCAGCCCTTTTAATCGACGATATAGAATAAT 540

Db 181 TCGCAAAATTTGAGCTCTTGCATAATTCAGCCCTTTTAATCGACGATATAGAATAAT 240

QY 541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCACTATA 600

Db 241 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCACTATA 300

QY 601 AACACCGCAAAATTTATCTATTTTCAGAGCCATGCACTTGTATCGCAGCTAACCAAAA 660

Db 301 AACACCGCAAAATTTATCTATTTTCAGAGCCATGCACTTGTATCGCAGCTAACCAAAA 360

QY 661 GAGCCTTTGTATCATTAATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCTACATAGG 720

Db 361 GAGCCTTTGTATCATTAATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCTACATAGG 420

QY 721 GGACAGGCTTGGATATATATCTGGAGAGACTTTCTGCTGTAATCATACCTTACTCAGGAG 780

Db 421 GGACAGGCTTGGATATATATCTGGAGAGACTTTCTGCTGTAATCATACCTTACTCAGGAG 480

QY 781 ATGTATTTGAATATCGTTTATGAATAAAACAGCGCGCTTTTCAGATTAACTGAGACTC 840

Db 481 ATGTATTTGAATATCGTTTATGAATAAAACAGCGCGCTTTTCAGATTAACTGAGACTC 540

QY 841 ATGGAAGCGCTGCTCTCCCTTCACACACCGCCCAATTCGTTGGTTCCTTCAATAATCTT 900

Db 541 ATGGAAGCGCTGCTCTCCCTTCACACACCGCCCAATTCGTTGGTTCCTTCAATAATCTT 600

QY 901 CTGGGTATTTATTTATCAGATTAGAGATGATTACTTTGAAATTTGAAAGATTTCCAAATGTC 960

Db 601 CTGGGTATTTATTTATCAGATTAGAGATGATTACTTTGAAATTTGAAAGATTTCCAAATGTC 660

QY 961 AGCGAAAAAGGCTTTGCTGAGACATTAACAGAGGGAAGTTATCTTTTCCCATCTGCCAC 1020

Db 661 AGCGAAAAAGGCTTTGCTGAGACATTAACAGAGGGAAGTTATCTTTTCCCATCTGCCAC 720

QY 1021 GCCCTTAACTTCACTAAAACGAAAGTCAAACTGAGCAACACAATGAAATTCCTAAGAAT 1080

Db 721 GCCCTTAACTTCACTAAAACGAAAGTCAAACTGAGCAACACAATGAAATTCCTAAGAAT 780

QY 1081 CTCCTGTTGAGGACAAGTGATAAAGATATAAAGCTTAAAGCTGATTCAATATCTGGAATTC 1140

Db 781 CTCCTGTTGAGGACAAGTGATAAAGATATAAAGCTTAAAGCTGATTCAATATCTGGAATTC 840

QY 1141 GACACCAATTCATTTGGCTTACACCAAAAATTTTATTAATCAATAGTGAATATGATAAAA 1200

Db 841 GACACCAATTCATTTGGCTTACACCAAAAATTTTATTAATCAATAGTGAATATGATAAAA 900

QY 1201 AATGATATGAAATAAGTATTTACCTGATTTGGCTTCGATTCGACACCGCCACCAAT 1260

Db 901 AATGATATGAAATAAGTATTTACCTGATTTGGCTTCGATTCGACACCGCCACCAAT 960

QY 1261 TTACATGACGAATTTGTTATATATATATATAGACCACTTATCCGAATTTGTA 1308

Db 961 TTACATGACGAATTTGTTATATATATATATAGACCACTTATCCGAATTTGTA 1008

RESULT 3

ABK96782

ID ABK96782 standard; DNA; 1008 BP.

XX AC ABK96782;

XX DT 24-SEP-2002 (first entry)

XX DE S. cerevisiae prenyl diphosphate synthase gene #2.

XX KW Prenyl alcohol; prenyl diphosphate synthase; geranylgeraniol;

XX KW hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; gene; ss;

XX KW isopentenyl diphosphate delta-isomerase; mevalonate kinase;

XX KW mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.

XX OS Saccharomyces cerevisiae.

XX PN WO200253746-A1.

XX XX 11-JUL-2002.

XX XX 20-DEC-2001; 2001WO-JP011214.

XX XX 28-DEC-2000; 2000JP-00403067.

XX XX (TOYT) TOYOTA JIDOSHA KK.

XX XX Ohto C, Obata S, Muramatsu M, Nishi K, Totusuka K;

XX XX WPI; 2002-537944/57.

XX Production of prenyl alcohols by culturing a transformant transferred

XX with e.g. prenyl diphosphate synthase gene, for use in industrial

XX synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds.

Example 2; Page 182-185; 335pp; Japanese.

The invention relates to a process for producing a prenyl alcohol comprising: (A) construction of a recombinant by transferring an expression recombinant DNA or a DNA for genome integration into a host which contains prenyl diphosphate synthase gene or its variant; and (B) collecting product from the culture medium. Also described is: (1) a method for producing prenyl alcohol in which the expression recombinant DNA contains: (a) a hydroxymethylglutaryl-CoA reductase gene or its variant; or (b) an isopentenyl diphosphate delta-isomerase gene; (2) a process for producing geranylgeraniol, comprising: (a) constructing a recombinant by transferring an expression recombinant DNA or a DNA for genome integration into a host which contains hydroxymethylglutaryl-CoA reductase gene or its variant; and (b) isolating the product; (3) a process for producing farnesylgeraniol in which the expression recombinant DNA also contains an isopentenyl diphosphate delta-isomerase gene, mevalonate CoA acetyltransferase gene, hydroxymethylglutaryl-CoA synthase gene, mevalonate kinase gene, or mevalonate diphosphate decarboxylase gene. The methods are used for the production of prenyl alcohols, particularly for use in industrial synthesis of isoprenoid-terpenoid compounds e.g. physiologically-active prenyl alcohols including geometric isomers. ABK96780-ABK96897 represent prenyl diphosphate synthase genes and related PCR primers used in the methods of the invention

Sequence 1008 BP; 337 A; 199 C; 176 G; 296 T; 0 U; 0 Other;

Query Match 64.2%; Score 1008; DB 6; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 9.4e-229;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

301 ATGAGGCCAAGATAGATGAGCTGATCAATAATGATCCCTGTTGGTCCAGCCAAAATGAA 360
 1 ATGAGGCCAAGATAGATGAGCTGATCAATAATGATCCCTGTTGGTCCAGCCAAAATGAA 60

361 AGCTTCAATTTCAAAACCTTATATATCATCATCTCTTTGAAACCTGGCAAGAACTTTAGACTA 420
 61 AGCTTCAATTTCAAAACCTTATATATCATCATCTCTTTGAAACCTGGCAAGAACTTTAGACTA 120

421 AATTAAATGTTCAAAATTAACAGAGTTATGAATTTGCCGCCAAGACGAGTGGCCATGTT 480
 121 AATTAAATGTTCAAAATTAACAGAGTTATGAATTTGCCGCCAAGACGAGTGGCCATGTT 180

481 TCGCAATTTGTTGAGCTCTTGCAATAATCCAGACCTTTTAAATCGACGATATAGAAAGATAAT 540
 181 TCGCAATTTGTTGAGCTCTTGCAATAATCCAGACCTTTTAAATCGACGATATAGAAAGATAAT 240

541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
 241 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 300

601 AACACGGCAATATATATGATTTTACAGAGCCATGCACTTGTATCGAGCTAACACAAAA 660
 301 AACACGGCAATATATATGATTTTACAGAGCCATGCACTTGTATCGAGCTAACACAAAA 360

661 GAGCCTTTGATCATATTTGATTTTACGATTTTCAACGAAGATTTGATCAATCTACATAGG 720
 361 GAGCCTTTGATCATATTTGATTTTACGATTTTCAACGAAGATTTGATCAATCTACATAGG 420

721 GGAACAAGCTTGGATATATCTGAGAGACTTTCTGCTCGAAATCATACCTACTCAGGAG 780
 421 GGAACAAGCTTGGATATATCTGAGAGACTTTCTGCTCGAAATCATACCTACTCAGGAG 480

781 ATGTATTTGAATGTTGATGATTAATAACAGCGGGCTTTTCAGATTAACGTTGAGACTC 840
 481 ATGTATTTGAATGTTGATGATTAATAACAGCGGGCTTTTCAGATTAACGTTGAGACTC 540

841 ATGGAAGCGTGTCTCTCTCTCTCACACAGCGCAATCTGTTGTTCTCTTTCAAAATCTT 900
 541 ATGGAAGCGTGTCTCTCTCTCTCACACAGCGCAATCTGTTGTTCTCTTTCAAAATCTT 600

901 CTGGGTATTTATTCAGATAGATGATTAATCTGAAATTTGAAAGATTTCCAAATGTCC 960

601 CTGGGTATTTATTTATCAGATTAGAGATTAGATTACTTTGAAATTTGAAAGATTTCCAAATGTCC 660

961 AGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCCAC 1020

661 AGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCCAC 720

1021 GCCCTTAACCTTCACTAAACGAAAGGTCAAACCTGAGCAACACATGAAATTTAAAGAAAT 1080

721 GCCCTTAACCTTCACTAAACGAAAGGTCAAACCTGAGCAACACATGAAATTTCTAAGAAAT 780

1081 CTCTGTTGAGCAAGTATGATTAAGATATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1140

781 CTCTGTTGAGCAAGTATGATTAAGATATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 840

1141 GACACCAATTCATGCGCTTACACCAAAATTTTATTAATCAATTAAGTAAATGATGATAAAA 1200

841 GACACCAATTCATGCGCTTACACCAAAATTTTATTAATCAATTAAGTAAATGATGATAAAA 900

1201 AATGATAATGAAATAAGTATTTACCTGATTTGGCTTCGCATTCGACACCGCCCAAT 1260

901 AATGATAATGAAATAAGTATTTACCTGATTTGGCTTCGCATTCGACACCGCCCAAT 960

1261 TTATGAGCAAGTATGATTAATATATATAGACCACTTATCCGAAATTTGA 1308

961 TTATGAGCAAGTATGATTAATATATATAGACCACTTATCCGAAATTTGA 1008

RESULT 4
 ADC49385
 ID ADC49385 standard; DNA; 1008 BP.
 XX
 AC ADC49385;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Yeast BTS1 cDNA.
 XX
 KW Yeast; ss; gene; microbiological; prenyl alcohol; geranylgeraniol;
 KW farnesol; nerolidol; squalene synthetase inhibitor; BTS1.
 XX
 OS Saccharomycetes cerevisiae.
 XX
 FH Key Location/Qualifiers
 FT 1. .1008
 FT /*tag= a
 FT /product= "BTS1"
 XX
 PN JP2002300896-A.
 XX
 PD 15-OCT-2002.
 XX
 PF 18-JAN-2002; 2002JP-00010528.
 XX
 PR 30-JAN-2001; 2001JP-00021547.
 XX
 PA (TOYT) TOYOTA JIDOSHA KK.
 XX
 WI WI; 2003-485780/46.
 DR P-PSDB; ADC49388.
 DR
 PT A process for preparation of prenyl alcohols by culture of their
 PT productive microorganisms in a medium with added squalene synthetase
 PT inhibitor.
 XX
 PS Disclosure; SEQ ID NO 12; 37pp; Japanese.
 XX
 CC The invention discloses a microbiological preparation of prenyl alcohols,
 CC particularly geranylgeraniol, farnesol and/or nerolidol, by culture of
 CC their productive microorganism spp. of Saccharomycetes, Saccharomycopsis,
 CC Saccharomycodes, Schizosaccharomycetes, Wickerhamia, Debaryomyces,
 CC Hanseniaspora, Pichia, Candida, Zygosaccharomycetes, Ogataea, Kurashia,
 CC Komagataella, Yarrowia, Kluyveromyces, Trichosporon, Cryptococcus,
 CC Torulaspora, Bacillus, Staphylococcus, Micrococcus, Exiguobacterium,

CC one selected from farnesyl diphosphate synthase gene, geranylgeranyl
 CC diphosphate synthase gene, hydroxymethylglutaryl-CoA reductase gene,
 CC isopentenyl diphosphate A-isomerase gene, mevalonate kinase gene, acetyl-
 CC CoA acetyltransferase gene, hydroxymethylglutaryl-CoA synthase gene,
 CC phosphomevalonate kinase gene and diphosphomevalonate decarboxylase gene,
 CC or a fusion gene composed of a gene selected from the above genes, or a
 CC gene obtained by introducing an addition, substitution or insertion
 CC mutation into any one of the above genes so that the polypeptide encoded
 CC by the resultant gene contains an endoplasmic reticulum signal. The
 CC methods and compositions of the present invention are useful for
 CC producing prenyl alcohols by reducing the amount of squalene synthase
 CC gene transcript having translational activity. The present sequence
 CC encodes an IPP (isoprenyl diphosphate) biosynthetic pathway-related
 CC enzyme.

XX SQ Sequence 1008 BP; 337 A; 199 C; 176 G; 236 T; 0 U; 0 Other;

Query Match 64.2%; Score 1008; DB 12; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 9, 4e-229;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATATGATCCTCTGTTGGTCCAGCCAAATGAA 360
 Db 1 ATGGAGGCCAAGATAGATGAGCTGATCAATATATGATCCTCTGTTGGTCCAGCCAAATGAA 60

QY 361 AGCTTGATTTCAAAACCTTATATACATCCTCTTTGAAACCTGGCAAGAACTTTAGACTA 420
 Db 61 AGCTTGATTTCAAAACCTTATATACATCCTCTTTGAAACCTGGCAAGAACTTTAGACTA 120

QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATGAT 480
 Db 121 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATGAT 180

QY 481 TGCATAATTTGAGCTCTTGCATTAATTCAGCCTTTTAAATGCAAGATATAGAGATAAT 540
 Db 181 TGCATAATTTGAGCTCTTGCATTAATTCAGCCTTTTAAATGCAAGATATAGAGATAAT 240

QY 541 GCTCCCTTCAGAGGGGACAGACCTCTCTCACTTAATCTTGGTGTCACCTCCACTATA 600
 Db 241 GCTCCCTTCAGAGGGGACAGACCTCTCTCACTTAATCTTGGTGTCACCTCCACTATA 300

QY 601 AACACCGCAAAATATATATATTTTTCAGAGCCATGCAACTTTGTATCGAGCTTAACCAAAA 660
 Db 301 AACACCGCAAAATATATATATTTTTCAGAGCCATGCAACTTTGTATCGAGCTTAACCAAAA 360

QY 661 GAGCCTTTGATCATTAATTTGATATACGATTTTCAACGAGAAATGATCACTACATAGG 720
 Db 361 GAGCCTTTGATCATTAATTTGATATACGATTTTCAACGAGAAATGATCACTACATAGG 420

QY 721 GGACAGGCTTGATATATATCTGAGAGACTTTCTGCTGAAATCATACCTACTCAGGAG 780
 Db 421 GGACAGGCTTGATATATATCTGAGAGACTTTCTGCTGAAATCATACCTACTCAGGAG 480

QY 781 ATGTATTTGAATATGTTATGAATAAAACAGCGGCGCTTTTCAGATTAACGTTGAGACT 840
 Db 481 ATGTATTTGAATATGTTATGAATAAAACAGCGGCGCTTTTCAGATTAACGTTGAGACT 540

QY 841 ATGGAAGCGTGCTCTCTCCTCACACCGGCGCATTCGTTGCTTTTCATAAATCTT 900
 Db 541 ATGGAAGCGTGCTCTCTCCTCACACCGGCGCATTCGTTGCTTTTCATAAATCTT 600

QY 901 CTGGGTATTTATTTATCAGATTTAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTC 960
 Db 601 CTGGGTATTTATTTATCAGATTTAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTC 660

QY 961 AGCGAAAAGCGCTTTGCTGAGGACATTAACAGAGGGAAGTTATCTTTCCATCGTCCAC 1020
 Db 661 AGCGAAAAGCGCTTTGCTGAGGACATTAACAGAGGGAAGTTATCTTTCCATCGTCCAC 720

QY 1021 GCCCTTAACCTCACTAAACGAAGGTCAACTGAGCAACACATGAATGAATTTCTAAGAAAT 1080
 Db 721 GCCCTTAACCTCACTAAACGAAGGTCAACTGAGCAACACATGAATGAATTTCTAAGAAAT 780

QY 1081 CTCCTGTTTGGAGCAAGTGATATAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC 1140
 Db 781 CTCCTGTTTGGAGCAAGTGATATAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC 840

QY 1141 GACACCAATTCATTTGGCCTACACCAAAATTTTATTAATCAATTAATGAATGATAAAA 1200
 Db 841 GACACCAATTCATTTGGCCTACACCAAAATTTTATTAATCAATTAATGAATGATAAAA 900

QY 1201 AATGATATGAAAAATAAGTATTTTACCTGATTTGGCTTCGCATTCGACACCCGCCCAAT 1260
 Db 901 AATGATATGAAAAATAAGTATTTTACCTGATTTGGCTTCGCATTCGACACCCGCCCAAT 960

QY 1261 TTACATGACGAATTTGTTATATATAATAGACCACTTATCCGAATTTGGA 1308
 Db 961 TTACATGACGAATTTGTTATATATAATAGACCACTTATCCGAATTTGGA 1008

RESULT 6

ADK52980

ID ADK52980 standard; DNA; 599 BP.

XX AC ADK52980;

XX DT 06-MAY-2004 (first entry)

XX DE Plant DNA sequence which confers altered metabolic characteristic #363.

XX KW altered metabolic characteristic; plant; acid metabolism;

XX KW alcohol metabolism; fatty acid metabolism; alkaloid metabolism;

XX KW branched fatty acid metabolism; ester metabolism; glyceride metabolism;

XX KW amino acid metabolism; carboxylate metabolism; steroid metabolism;

XX KW phenolic metabolism; carbohydrate metabolism; alkene metabolism;

XX KW terpene metabolism; isoprenoid metabolism; ketone metabolism;

XX KW alkyne metabolism; hydrocarbon metabolism; disease resistance; gene shuffling; sexual PCR; ds.

XX KW quinone metabolism; disease resistance; gene shuffling; sexual PCR; ds.

XX OS Saccharomycetes cerevisiae.

XX PN WO2003020936-A1.

XX PD 13-MAR-2003.

XX PF 30-AUG-2002; 2002WO-US027884.

XX PR 31-AUG-2001; 2001US-0316471P.

XX PA (DOWC) DOW CHEM CO.

XX PA (DOWC) DOW AGROSCIENCES LLC.

XX PI Weglarz T, Gachotte D, Blakeslee B, McCreary DA, Pell RJ;

XX PI Oriedo JVB, Crosley R, Reddy AS, Shukla V, Larrinua I, Miller BA;

XX DR WPI; 2003-313091/30.

XX CC Novel genes that confer altered metabolic characteristics in Nicotiana benthamiana plants, useful for altering the levels of metabolites e.g. acids, fatty acids, amino acids, carbohydrates, hydrocarbons and sterols.

XX PS Claim 1; SEQ ID NO 363; 2576pp; English.

XX CC The invention comprises DNA sequences which confer an altered metabolic characteristic when they are expressed in a plant. The DNA sequences of the invention are useful for producing plants with an altered metabolic characteristic, such as: altered acid metabolism, alcohol metabolism, fatty acid metabolism, branched fatty acid metabolism, alkaloid or other base metabolism, altered amino acid metabolism, altered ester metabolism, altered glyceride metabolism, altered phenolic metabolism, altered or CC carboxylate metabolism, altered steroid, oxygenated terpene, or CC isoprenoid metabolism, alkene or alkyne metabolism, hydrocarbon CC metabolism, ketone or quinone metabolism. The DNA sequences of the CC invention may be used to provide disease resistance in a plant and gene CC shuffling or sexual PCR procedures. The present nucleic acid represents a CC DNA sequence of the invention.

XX SQ Sequence 599 BP; 193 A; 129 C; 106 G; 171 T; 0 U; 0 Other;
 Query Match 36.2%; Score 568.6; DB 10; Length 599;
 Best Local Similarity 98.5%; Pred. No. 1e-124;
 Matches 574; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 299 CAATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCAGCCAAAATG 358
 Db 17 CCATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCAGCCAAAATG 76
 QY 359 AAAGCTTGATTTTAAAAACCTTAAATCAATCAATCTTTTGAACCTGGCAAGAACTTTAGAC 418
 Db 77 AAAGCTTGATTTTAAAAACCTTAAATCAATCAATCTTTTGAACCTGGCAAGAACTTTAGAC 136
 QY 419 TAAATTTAATAGTTCAATTAACAGAGTATGATTTTGGCCAAAGACAGCTGGCCATAG 478
 Db 137 TAAATTTAATAGTTCAATTAACAGAGTATGATTTTGGCCAAAGACAGCTGGCCATAG 196
 QY 479 TTTGCGAAATGTTGAGCTCTTGATATATCCAGCTTTTAAATCGAGATATAGAGATA 538
 Db 197 TTTGCGAAATGTTGAGCTCTTGATATATCCAGCTTTTAAATCGAGATATAGAGATA 256
 QY 539 ATGCTCCCTTGAGAAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTTCCACTA 598
 Db 257 ATGCTCCCTTGAGAAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTTCCACTA 316
 QY 599 TAAACACCGCAATATATGATTTTCAAGGCGATGCAACTGTATCGCAGCTTAACCAAA 658
 Db 317 TAAACACCGCAATATATGATTTTCAAGGCGATGCAACTGTATCGCAGCTTAACCAAA 376
 QY 659 AAGAGCCTTTGTTATCATATTAATTTGATTAACGATTTTCAACGAAGAAATTCATCAATCTACATA 718
 Db 377 AAGAGCCTTTGTTATCATATTAATTTGATTAACGATTTTCAACGAAGAAATTCATCAATCTACATA 436
 QY 719 GGGGACAAGGCTTGGATATATACATGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778
 Db 437 GGGGACAAGGCTTGGATATATACATGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 496
 QY 779 AGATGATTTTGAATATGTTTGAATTAATTAACAGAGGCGGCTTTTCAAGATTAACGTTGAGAC 838
 Db 497 AGATGATTTTGAATATGTTTGAATTAATTAACAGAGGCGGCTTTTCAAGATTAACGTTGAGAC 556
 QY 839 TCATGGAAGCGCTGCT 881
 Db 557 TCATGGAACAACGCT 599

RESULT 7
 ADF75080
 ADF75080 standard; DNA; 963 BP.
 AC ADF75080;
 XX
 DT 26-FEB-2004 (first entry)
 XX
 DE A_gossypii DNA essential for fungal growth & development SeqID 21.
 XX
 KW fungal growth; fungicide; screening assay; fungicidal; crop; gene; ds.
 XX
 OS Bremothecium gossypii.
 XX
 PN W02003072706-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 11-FEB-2003; 2003WO-US004154.
 XX
 PR 22-FEB-2002; 2002US-0359557P.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Gaffney TD, Flavien A, Dietrich F, Philippsen P, Kirksey CMW;

PI Gates K, Gardner J;
 XX WPI; 2003-721756/68.
 DR P-PSDB; ADF75081.
 XX
 PT Identifying a fungicidal compound comprising combining a polypeptide with
 PT the compound to be tested with the ability to bind, or to inhibit the
 PT activity of the polypeptide under conditions conducive to binding or
 PT inhibition.
 XX
 PS Claim 11; SEQ ID NO 21; 127pp; English.
 XX
 CC This invention relates to novel nucleic acid molecules, isolated from
 CC Aeshya gossypii, which encode proteins essential for fungal growth and
 CC development. Specifically, it refers to methods for identifying these
 CC proteins as fungicide targets by using gene disruption with a genetic
 CC resistance gene. In particular, the nucleic acids encoding the proteins
 CC required for fungal growth provide the basis of screening assays designed
 CC to easily and rapidly identify inhibitory compounds that exhibit
 CC fungicidal activity. Accordingly, the present invention describes such
 CC methods to identify protein inhibitors that can be used as fungicides to
 CC suppress the growth of undesirable fungi in agronomically important crops
 CC such as maize, wheat, barley, cotton and sugar beet. This polynucleotide
 CC sequence is a DNA sequence essential for fungal viability, used in an
 CC exemplification of the invention.
 XX
 SQ Sequence 963 BP; 206 A; 252 C; 293 G; 212 T; 0 U; 0 Other;
 Query Match 11.7%; Score 183.4; DB 10; Length 963;
 Best Local Similarity 54.7%; Pred. No. 2.2e-33;
 Matches 458; Conservative 0; Mismatches 361; Indels 18; Gaps 4;
 QY 374 AACCTTAAATCAATCCTTTTGAACCTGGCAAGAACTTTAGACTAAATTTAATAGTTC 433
 Db 71 AGCTTTTATATACCTATAGTACGCGCGGCAAGAAATTTCCGCAATACGCTGATCGCG 130
 QY 434 AAATTAACAGATTAATGAATTTGCCAAAGACAGCTGGCCATAGTTTCGCAATTTCTG 493
 Db 131 TGTTCAACGGGTTCTATGTTCTTTCGAGCGGAGGTCGCGCGGTCAAGAGTTGTGTG 190
 QY 494 AGCTTTTGCATAAATCCAGCTTTTAAATCGACGATATAGAGATATATGCTCCCTTGAGAA 553
 Db 191 AGATGTTGATGCGGAGCTGTTGATCGACGATATCGAAGACAACTCTGCTAGGAGAC 250
 QY 554 GGGGACAGACCACTTCTCACTTAATCTTCTGTTGATACCTCCATATAAACAACGCAAT 613
 Db 251 GTGCGCTGCTGCTGCGCCATGATATGCTGCGCCATGAGATCAACACTGCGAAT 310
 QY 614 ATATGATTTTCAGAGCATCAACTTTGATCGCAGCTAAACCAAAA-----AGAGCCTT 667
 Db 311 ACATGATTTTGTCTCAATGCTGCTGCTGCGGAGCTGCGCGGTGAGCGTCCGCGGGGCC 370
 QY 668 TGATCATATTTGATTAACGATTTTCAACGAAGAAATTTGATCAATCTACATAGGGGACAAG 727
 Db 371 CGCTGCGAGGATCTGCTGAAGTCTTCAACGAGGAGATGATGATCTGCACTCGCGGGCAG 430
 QY 728 GCTTGATATATATCTGAGAGACTTTCTGCTGAAATATACATACCTACTCAGAGAGATGAT 787
 Db 431 GCCTTGACATATATTTGGCGCGACACGTTTACGG-----TGCGTGGGAGCATCACTACC 484
 QY 788 TGAATATGTTATGATAAAACAGCGCGCTTTTTCAGATTAACGTTGAGACTCATGGAAG 847
 Db 485 TCGGATGTTGATGATATAGACCGCGCGCTTTGTTCCGGCTGACCGGTGCGGTATCATGGAAG 544
 QY 848 CGCTGTCTCCTCTCACA---CCACGGCCATTTCTGTTGTTCTCTTTTCATATAATCTTCTGG 904
 Db 545 CGTTACGGGAGGCGCCAGATGGACCCGGATCAACATTTGTTGCGCTTAGCAACTGCTTG 604
 QY 905 GTATTTATTTATCAGATTAGAGATGATTAATTTGAAATTTGAAATTTCCAAATTTCCAGCG 964
 Db 605 GCGTGTGTACCAAGGTGCGGAGACTATTTTGAACCTGACCGCACTCGCGCATTCGGA 664
 QY 965 AAAAGGCTTTGCTGAGGACATTAACAGAGGGAAGTTATCTTTTCCATCGTCCAGCCC 1024

```

665 ACAAGGCTTTGAGACGATATCATCCGAGGCAAAATTCCTACCCATTATCCAGGCG 724
1025 TTAACCTCACTAAACGAAAGTCAAACTGAGCAACACAAATGAATTCATGCTCC 1084
725 TGCAGTACGCGGTGTCCA---TGACCCAGCGGCTACGACTTCCTGGTCACTGTGTAC 781
1085 TGTGAGACAGTCAATGAATATAAATAAAGCTGATTAATCAATATCTGGAATTCGACA 1144
782 GCGAGACGACAGATATCACCAAGCGCGGTGGATATCTTGGGACGTCA 841
1145 CCAATTCATTTGGCTTACCAACCAAAATTTTATTAATCAATTAATGATGATAAAA 1201
842 GTGGTTCGCTCGGTACAGGAGCAGCTATCATCGAAGCTATGCAACCAATGATAAGA 898

```

RESULT 8

AAT31846
ID AAT31846 standard; cDNA; 903 BP.

XX AC AAT31846;

XX DT 17-FEB-1997 (first entry)

XX DE Human geranylgeranyl pyrophosphate synthetase (GGPPS) cDNA sequence.

XX KW Geranylgeranyl pyrophosphate synthetase; cholesterol metabolism; steroid;
XX post-translational modification; gene therapy; rhabdomyolysis;
XX cell morphology; neoplasia; cancer; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX CDS 1..903
XX FT /*tag= a

XX PN WO9621736-A1.

XX PD 18-JUL-1996.

XX PF 11-JAN-1995; 95WO-US000421.

XX PR 11-JAN-1995; 95WO-US000421.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Greene JM, Rosen CA, Kirkness EP;

XX PS WPI; 1996-342289/34.

XX DR P-PSDB; AAR97565.

XX PT Polynucleotide encoding human geranylgeranyl pyrophosphate synthetase -
XX fragments of which can be used as probes to diagnose disease or
XX susceptibility to disease related to mutation in above polynucleotide.

XX PS Claim 7; Page 38; 55pp; English.

XX CC AAT31846 is a cDNA sequence encoding human geranylgeranyl pyrophosphate
XX synthetase (GGPPS). The sequence was obtained from a foetal heart cDNA
XX library and identified by homology to the corresponding enzyme of
XX Neurospora crassa. The GGPPS enzyme is involved in post-translational
XX modification (more specifically the geranylgeranylation) of proteins and
XX is also involved in a branch of the cholesterol/steroid metabolic
XX pathway. GGPPS catalyzes the trans-addition of 3 molecules of isopentenyl
XX diphosphate onto dimethylallyl diphosphate to form C20 GGPP. GGPP
XX controls the signal transduction, activation and trafficking of GGPP-
XX modified proteins. Human GGPPS can be used to control cell morphology
XX e.g. to treat rhabdomyolysis (a side effect of treatment with HMG-CoA
XX reductase) and unwanted apoptosis. Antagonists of GGPPS can be used to
XX treat neoplasms and to prevent viral proliferation, increased levels of
XX GGPPS are indicative of cancer cell growth

XX SQ Sequence 903 BP; 305 A; 173 C; 183 G; 242 T; 0 U; 0 Other;

```

Query Match          7.3%; Score 114; DB 2; Length 903;
Best Local Similarity 52.1%; Pred. No. 6.1e-17;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAAGCTTGATTTCAAAACCTTATATACATCCCTTTTGAACCTGGCAAGAACTTTAGAC 418
DB 26 AAAGAAATCTTCTAGAACCTTATAAATACTTACTTTCAGTTTACCAGGTAAACAAGTGAGAA 85
QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAG 478
DB 86 CCAAACTTTTCACAGGCATTTAATCATTTGGCTGAAAGTTCCAGAGGACAAGCTACAGATA 145
QY 479 TTTCGCAAAATTTGTTGAGCTCTTTGCATATAATTCAGAGCTTTTAAATCGACGATATAGAAGATA 538
DB 146 TTATTTGAAGTGACAGAAATGTTGCATATGCGAGTTTACTCATCGATGATTTGAAGACA 205
QY 539 ATGCTCCCTTGAGAAAGGGGACAGACCACTTCTACATTAATCTTCGGTGTACCTCCACTA 598
DB 206 ACTCAAAACTCCGACGCTGGCTTCCAGTGGCCACACAGCATCTATGGAATCCCATCTGTCA 265
QY 599 TAAACACCGCAAAATTAATATGATATTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAA 658
DB 266 TCAATCTGCCAATTAACGTTATTTCTTGGCTTGGAGAAAGTCT----- 310
QY 659 AAGAGCCTTTGTATCATATAATTTGATTCAGATTTTCAACGAAAGATTTGATCAATCTACATA 718
DB 311 TAAACCTTTGATCACCAGATGCAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 370
QY 719 GGGGCAAGGCTTGATATATATCTGAGAGACTTCTCGCTGGAATCATACCTACTCAGG 778
DB 371 AGGCAAGGCTCTAGATATTTTACTGGAGGGA-----TAAATTACACTTTGTCCCACTGAAG 424
QY 779 AGATGTTATTTGAATATGTTATGAATAAAACAGGGCGCTTTTCAGATTAAAGTTTGAGAC 838
DB 425 AAGAATATAAGCTATGTTGCTGCAGAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC 484
QY 839 TCATGGAAGCGCTGTCTCCTTCTCACACCGGCGCATTCGTTGGTTCCTTTTCATAAAATC 898
DB 485 TCATGCACTGTCTCTGATTACAAAGAAAGATTTAAACCGCTACTT-----ATA 535
QY 899 TTCTGGGTATTTATTTATCAGATTAGAGATGATTTCAATTTGAAAGATTTCCAAATGT 958
DB 536 CACTTGGGCTCTTTTTCAAAATTAGGATGATTTATGCTAATCTACACTCCCAAGAAATATA 595
QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTCAGAGGGGAAGTTATCTTTTCCCATCGTCC 1018
DB 596 GTGAAAACAAAAGTTTGGTGAAGATCTGACAGAGGGAAGTTCTCATTTCTCTACTATTC 655
QY 1019 ACGCCCTT 1026
DB 656 ATGCTATT 663

RESULT 9
AAV41555
ID AAV41555 standard; cDNA; 903 BP.
XX AC AAV41555;
XX DT 12-OCT-1998 (first entry)
XX DE Human geranylgeranyl pyrophosphate synthase.
XX KW Human; geranylgeranyl pyrophosphate synthase; hGGPS; antagonist;
XX agonist; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..903
XX FT /*tag= a
XX PT /product= "human geranylgeranyl pyrophosphate synthase"

```

```

XX FN US5786193-A.
XX PD 28-JUL-1998.
XX PF 06-JUN-1995; 95US-00469665.
XX PR 06-JUN-1995; 95US-00469665.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Greene JM, Kirkness EF;
XX DR WPI; 1998-436529/37.
XX DR P-PSDB; AAW59753.
XX PT DNA encoding human geranyl pyrophosphate synthetase polypeptide -
XX PS useful for producing recombinant polypeptide.
XX CC Claim 5; Fig 1; 22pp; English.
XX CC This is the nucleotide sequence of the human geranyl geranyl
XX CC pyrophosphate synthase (hGPPS) used in the method of the invention. The
XX CC proteins are used in the production of recombinant hGPPS, which can be
XX CC used to screen for hGPPS antagonists or agonists
XX SQ Sequence 903 BP; 304 A; 173 C; 184 G; 242 T; 0 U; 0 Other;

Query Match 7.3%; Score 114; DB 2; Length 903;
Best Local Similarity 52.1%; Pred. No. 6.1e-17;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAAAGCTTGATTCAAAACCTTATATACATCACCCTTTTGAACCTGGCAAGACTTTAGAC 418
DB 26 AAGAATCTCTTAGAACCTTATATATCTTCTCAGTTACCAAGGTAACAGTGAGAA 85
QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGCCCATAG 478
DB 86 CCAAACTTTTCAAGGCAATTTAATCATTTGGCTGGAAGTTCCAGAGGACAAGCTACAGATTA 145
QY 479 TTTGCGAAATTTGAGCTCTTCATATTTCCAGCCCTTTTAAATCGACGATATAGAGATA 538
DB 146 TTATTGAAGTGACAGAAATTTGTCATAATGCCAGTTTACTCATCGATGATATTGAAGACA 205
QY 539 ATGCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTA 598
DB 206 ACTCAAACTCCGACGTGGCTTCCAGTGGCCCAAGCACTATGGAATCCCATCTGTCA 265
QY 599 TAAACACCGCAAAATATATATGTTATTCAGAGCCATGCACTTGTATCGCAGCTAACCAAA 658
DB 266 TCAATTCGCAATTTACGTGTATTTCTTGGCTTGGAGAAAGTCT----- 310
QY 659 AAGAGCCTTTGTATCATATTTGATTTAGATTTTCAACGGAAGATTTGATCAATCTACATA 718
DB 311 TAACCCCTTGATCACCCAGATGCACTGAAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 370
QY 719 GGGGACAAGGCTTGGATATATCTAGAGAGACTTCTGCTGGAATCATACCTACTACTAGG 778
DB 371 AGGACAAGGCTAGATATTTTACTTGGAGGGA-----TAAATACATTTGCCACTGAAG 424
QY 779 AGATGTTATTTGAATGTTATGTAATAAAACAGCGGCGCTTTTCAGATTAACAGTTGAGAC 838
DB 425 AAGAATATAAGCTATGTTGCTGAGAAACAGGTGGACTGTTTGGATTAGCAGTAGTC 484
QY 839 TCATGGAAGCGCTGCTCTCTCCCTCAACCAACGCGCCATTCGTTGGTTCCCTTTCAATAATC 898
DB 485 TCATGCAAGTTGTTCTCTGATTAACAAAGAGATTTTAAACCGCTACTT-----AATA 535
QY 899 TTTCTGGCTATTTATATCAGATTTAGATGATTTACTTGAATTTGAAGATTTCCAAATGT 958
DB 536 CACTTGGGCTCTTTTTCCAAATTAGGATGATTTATGCTTAATCTACACTCCAAAGATATA 595
QY 959 CCAGCGAAAAAGGCTTTGCTGAGGACATTAACAGAGGGGAAGTTATCTTTTCCCATCTGCC 1018

```

```

DB 596 GTGAAAACAAAGTTTGGGTGAAGATCTGACAGAGGGGAAGTTCTCATTTCTACTATTC 655
QY 1019 AGCCCTT 1026
DB 656 ATGCTATT 663

RESULT 10
AAX86773
ID AAX86773 standard; cDNA; 903 BP.
XX AC AAX86773;
XX DT 20-SEP-1999 (first entry)
XX DE Human geranylgeranyl pyrophosphate synthetase (hGPPS) encoding cDNA.
XX KW Geranylgeranyl pyrophosphate synthetase; hGPPS; cell morphology; human;
XX KW apoptosis; neoplasia; tumour; cancer; viral proliferation; gene therapy;
XX KW human Choroideremia; ss.
XX OS Homo sapiens.
XX FN US928924-A.
XX PD 27-JUL-1999.
XX PF 11-MAR-1998; 98US-00038596.
XX PR 11-JAN-1995; 95WO-US000421.
XX PR 06-JUN-1995; 95US-00469665.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Kirkness EF, Greene JM;
XX DR WPI; 1999-429504/36.
XX DR P-PSDB; AAY27005.
XX PT Geranylgeranyl pyrophosphate synthetase peptides useful for regulating
XX CC apoptosis.
XX PS Claim 18; Fig 1A-D; 23pp; English.
XX CC This cDNA encodes a human geranylgeranyl pyrophosphate synthetase (hGPPS)
XX CC polypeptide. The hGPPS polynucleotides and polypeptides may be used to
XX CC control the morphology of cells, suppress apoptosis, screen for hGPPS
XX CC agonists and antagonists (which may in turn be used to treat neoplasia,
XX CC e.g. tumours and cancer cell growth, and prevent viral proliferation),
XX CC gene therapy, and diagnose a disease or susceptibility to diseases
XX CC related to a mutation in hGPPS (e.g. human choroideremia)
XX SQ Sequence 903 BP; 304 A; 173 C; 184 G; 242 T; 0 U; 0 Other;

Query Match 7.3%; Score 114; DB 2; Length 903;
Best Local Similarity 52.1%; Pred. No. 6.1e-17;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAAAGCTTGATTCAAAACCTTATATCACAATCTTTTGAACCTGGCAAGACTTTAGAC 418
DB 26 AAGAATCTCTTAGAACCTTATATATCTTCTCAGTTACCAAGGTAACAGTGAGAA 85
QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGCCCATAG 478
DB 86 CCAAACTTTTCAAGGCAATTTAATCATTTGGCTGGAAGTTCCAGAGGACAAGCTACAGATTA 145
QY 479 TTTGCGAAATTTGAGCTCTTCATATTTCCAGCCCTTTTAAATCGACGATATAGAGATA 538
DB 146 TTATTGAAGTGACAGAAATTTGTCATAATGCCAGTTTACTCATCGATGATATTGAAGACA 205
QY 539 ATGCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTA 598

```

Db 206 ACTCAAACTCCGAGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 265
QY 599 TAACACACGCAAAATTATATGATTTTCAGAGCCATCAACTGTGTATCGCAGCTTAACACAA 658
Db 266 TCAATTTCTGCCAATTAACGTGTATTTCTTGGCTTGGAGAAAGTCT----- 310
QY 659 AAGAGCCTTTGTATCATATTTGATTACGATTTTCAACGAAGAATTGATCAATCTACATA 718
Db 311 TAACCTTTGATCACCAGATGAGTGAAGCTTTTACCGCAGCTTTTGGAACTCCATC 370
QY 719 GGGGACAGGCTTGTATATATCTACTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGG 778
Db 371 AGGACAAAGGCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 424
QY 779 AGATGATTTGAATATGTTATGATTAATAAACAAGGCGCTTTTCAGATTAACGTTGAGAC 838
Db 425 AAGAATATAAAGCTATGTTGCTGCAGAAACAGAGTGGACTGTTTGGATTAGCAGTAGGTC 484
QY 839 TCATGGAAGCGCTGTCTCCTCTCACACACGCGCATTCGTTGGTTCCTTTTCATAAATC 898
Db 485 TCATGCAAGTGTCTCTGATTAACAAGAGATTTAAACCGCTACTT-----AATA 535
QY 899 TTCTGGGTATTTATCATAGATTAGAGATGATTACTTGAATTTGAAAGATTTTCCAAATGT 958
Db 536 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTAATCTACACTCCAAAGAATATA 595
QY 959 CCAGCGAAAAGCTTTGCTGAGGACATTTACAGAGGGAGTATCTTTTCCCATCGTCC 1018
Db 596 GTGAAAACAAAAGTTTGGGTGAGATCTGCACAGAGGAAAGTTCTCATTTCTACTATTTC 655
QY 1019 AGCGCCTT 1026
Db 656 ATGCTATT 663

RESULT 11
ADJ10255
ID ADJ10255 standard; DNA; 1395 BP.
XX AC ADJ10255;
XX 17-JUN-2004 (first entry)
XX Human geranylgeranyl diphosphate synthase 1 DNA SeqID 4.
XX gene; ds; chromosome 1q43; human; geranylgeranyl diphosphate synthase 1;
KW GGPS1; geranylgeranyl pyrophosphate synthetase; GGPS; ggppsase;
KW geranyltransferase; embryonic development; cell differentiation;
KW apoptosis; developmental; hyperproliferative disorder; cancer;
KW cytostatic.
XX Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 170..1072
XX /*tag= a
XX /*product= "GGPS1 protein"
XX US2004005570-A1.
XX PN
XX PD 08-JAN-2004.
XX PF 02-JUL-2002; 2002US-00189268.
XX PR 02-JUL-2002; 2002US-00189268.
XX (ISIS-) ISIS PHARM INC.
XX Dean NM, Bennett CF, Dobie KW;
XX WPI; 2004-201146/19.
XX P-PSDB; ADJ10383.
XX

PT New antisense oligonucleotides for modulating geranylgeranyl diphosphate
PT synthase 1 expression, useful for diagnosing, preventing or treating
PT conditions associated with the protein, e.g. cancer.
XX
XX Example 13; SEQ ID NO 4; 76pp; English.
XX This invention relates to a novel antisense compounds that modulate the
CC expression of human geranylgeranyl diphosphate synthase 1 (also known as
CC GGPS1, geranylgeranyl pyrophosphate synthetase, GGPS, ggppsase and
CC geranyltransferase) and located on chromosome 1p43. Specifically, it
CC refers to compositions useful for inhibiting the expression of GGPS1,
CC which normally participates in embryonic development, cell
CC differentiation and stimulates apoptosis via caspase-3 activation. The
CC present invention describes antisense oligonucleotides that comprise at
CC least one modified sugar moiety, a 2'-O-methoxyethyl (2' MOE) and at
CC least one modified nucleobase, a 5-methylcytosine. Accordingly, these
CC compounds are useful for treating a disease or condition associated with
CC GGPS1 such as a developmental or hyperproliferative disorder (e.g.
CC cancer) that arise as a result of aberrant apoptosis. As such, these
CC compositions exhibit cytostatic activity and are useful for diagnostics,
CC prophylaxis, research reagents and various kits. This polynucleotide
CC sequence is the geranylgeranyl diphosphate synthase 1 DNA of the
CC invention.
XX
SQ Sequence 1395 BP; 449 A; 259 C; 286 G; 401 T; 0 U; 0 Other;
Query Match 7.3%; Score 114; DB 12; Length 1395;
Best Local Similarity 52.1%; Pred. No. 7e-17;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
QY 359 AAAGCTTGTATTTCAAAACCTTATAATCACATCTTTTGAACCTCGCAAGACTTTAGAC 418
Db 195 AAAGAATTTCTTAGAACCTTATAATAATCTTACTTACAGTTACAGTAACAAGTGAGAA 254
QY 419 TAAATTTAAAGTTCAAAATTAACAGAGTTATGAAATTTGCCAAAGACGCTGGCCATAG 478
Db 255 CCAAACTTTTCACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACAGCTACGATTA 314
QY 479 TTTTCGCAATTTGTCAGCTTTGCGATAATTCAGAGCTTTTAATCGACGATATAGAGATA 538
Db 315 TTATTGAAGTGACAGAAATGTTGCAATATGCCAGTTTACTCATCATGATATATGAAGACA 374
QY 539 ATGCTCCCTTTGAGAAGGGGACAGACCCTTCTCATTAACTTCGGTGTACCTCCACTA 598
Db 375 ACTCAAACTTCGACGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 434
QY 599 TAAACACGCAAAATTAATGATTTTCAGAGCCATGCACTTGTATCGAGCTTAACACAA 658
Db 435 TCAATTTGCGCAATTAACGTGATTTTCTTGGCTTGGAGAAAGTCT----- 479
QY 659 AAGAGCCTTTGTATCATATTTGATTACGATTTTCAACGAAGAATTGATCAATCTACATA 718
Db 480 TAACCTTTGATCACCAGATGAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 539
QY 719 GGGGACAGGCTTGGATATATATCTACTGGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778
Db 540 AGGACACAGGCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 593
QY 779 AGATGATTTGAAATATGTTATGATTAATAAACAAGGCGCTTTTCAGATTAACGTTGAGAC 838
Db 594 AAGAATATAAAGCTATGTTGCTGCAGAAAACAGGTTGACTGTTGGATAGCAGTAGGTC 653
QY 839 TCATGGAAGCGCTGTCTCCTCTCTCACACACGCGCATTCGTTGGTTCCTTTTCATAAATC 898
Db 654 TCATGCAAGTGTCTCTGATTTACAAGAGATTTAAACCGCTACTT-----AATA 704
QY 899 TTCTGGGTATTTATCATAGATTAGAGATGATTTGATTTGAAAGATTTTCCAAATGT 958
Db 705 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTAATCTACACTCCAAAGAATATA 764
QY 959 CCAGCGAAAAGCTTTGCTGAGGACATTTACAGAGGGAGTATCTTTTCCCATCGTCC 1018
Db 765 GTGAAAACAAAAGTTTCTTGTGAAGATCTGCACAGAGGAAAGTTCTCATTTCTACTATTTC 824

QY 1019 ACGCCCTT 1026
 DB 825 ATGCTATT 832

RESULT 12
 AAZ60889
 ID AAZ60889 standard; cDNA; 1414 BP.
 XX
 AC AAZ60889;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE cDNA encoding a geranylgeranyl pyrophosphate synthetase (hGGPPS).
 XX
 KW Human; geranylgeranyl pyrophosphate synthetase; hGGPPS; chromosome 1;
 KW 1q42-1q43 locus; prostate cancer; hGGPPS; biallelic marker;
 KW mevalonic biosynthetic pathway; ss.
 XX
 OS Homo sapiens.
 XX

Key Location/Qualifiers
 CDS 85..987
 FT /*tag= a
 FT /product= "geranylgeranyl pyrophosphate synthetase"
 FT polyA_signal 1289..1294
 FT /*tag= b

WO200005382-A2.
 XX
 XX 03-FEB-2000.
 XX
 XX 23-JUL-1999; 99WO-1B001353.
 XX
 XX 23-JUL-1998; 98US-0093940P.
 XX
 XX (GEST) GENSET.
 XX
 XX Bougueleret L;
 XX
 XX WPI; 2000-182704/16.
 XX
 XX P-PSDB; AAY68909.
 XX
 PT New isolated human geranyl-geranyl pyrophosphate synthetase nucleic
 PT acids, used to develop agents for the diagnosis of, e.g. pathologies
 PT related to a defect in the mevalonic biosynthetic pathway.
 XX
 XX Claim 2; Page 80-82; 88pp; English.
 XX
 CC The present sequence encodes a human geranylgeranyl pyrophosphate
 CC synthetase (hGGPPS). Two differently spliced mRNAs exist for this gene.
 CC The first spliced mRNA is derived from a cDNA (AAZ60888) which comprises
 CC exons 1, 2, 3 and 4. The second mRNA is derived from a cDNA (AAZ60889)
 CC which comprises 1b1, 2, 3, and 4. The hGGPPS gene is located on
 CC chromosome 1, at the 1q42-1q43 locus. This chromosome 1 locus has been
 CC shown to carry a predisposing gene for prostate cancer. The nucleic acids
 CC encoding hGGPPS can be used for screening for agents which modulate the
 CC expression of the hGGPPS gene. Such agents can be used in therapeutic
 CC applications. The biallelic markers associated with the hGGPPS gene can
 CC be used for the diagnosis of diseases related to an alteration in the
 CC regulatory or coding regions of hGGPPS, such as pathologies related to a
 CC defect in the mevalonic biosynthetic pathway. The products can also be
 CC used for detection, diagnosis and drug screening
 XX
 XX SQ Sequence 1414 BP; 447 A; 284 C; 269 G; 409 T; 0 U; 5 Other;
 Query Match 7.3%; Score 114; DB 3; Length 1414;
 Best Local Similarity 52.1%; Pred. No. 7e-17;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
 359 AAAGCTGATTTCAAAACCTTATATACATCCTCTTTGAAACCTGGCAAGAACTTTAGAC 418
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 110 AAGAAATTTCTTAGAAACCCCTATATAATCTTACTTTCAGTTACCAAGTAAACAAGTGAGAA 169
 QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGAACCCAGCTGGCCATAG 478
 DB 170 CCAAACTTTTCACAGGCATTTAATCAATTTGGCTGAAAGTTCCAGAGGACAAAGCTACAGATTA 229
 QY 479 TTTCCGAAATTTGTTGAGCTCTTGTCATAAATTCACGCCCTTTTAATCGACGATATAGAAGATA 538
 DB 230 TTATTGAAGTGACAGAAATGTTGATTAATGCCAGTTTACTCATCGATGATATTAGAGACA 289
 QY 539 ATGCTCCCTTTGAGAAAGGGACAGACCACTTCTCTTAATCTTCGGTGTACCCCTCCACTA 598
 DB 290 ACTCAAACTCCGACGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 349
 QY 599 TAAACACCGCAAAATTAATGATTTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCAA 658
 DB 350 TCAATCTGCCAATTAAGTATTTCTTCTTGGCTTTGGAGAAAGTCT----- 394
 QY 659 AAGAGCCCTTTGTATCATAAATTTGATTACGATTTTCAACGAAGAATTTGATCAATCTACATA 718
 DB 395 TAACCTTTGATCACCAGATGCAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 454
 QY 719 GGGACAAAGCTTGGATATATATCTGAGAGACATTTCTGCTGAAATCATACCTACTCAGG 778
 DB 455 AGGACAAAGCCTAGATATTTTACTTGGAGGA-----TAATTAACACTTGTCCCACTGAAG 508
 QY 779 AGATGATTTGAATATGTTTATGAATAAAACAGCGCGCTTTTCAGATTAACGTTGAGAC 838
 DB 509 AAGAAATATAAAGCTATGCTGCTGCAGAAAAACAGGTGGACTGTTTGGATTAGCAGTAGTTC 568
 QY 839 TCATGGAAGCGCTGCTCTCTCTTCACACCAAGCCCATTCGTTGGTTCCCTTTCATAAATC 898
 DB 569 TCATGCAAGTTGTTCTCTGATTACAAGAAGATTTAAACCGCTACTTT-----ATA 619
 QY 899 TTTGGGTATTTATTTATCAGATTAGACATGATTACTTTGAATTTGAAAGATTTCCAAATGT 958
 DB 620 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTAATCTACTACTCCAAAGAATATA 679
 QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTCCCATCTGCC 1018
 DB 680 GTGAAAAACAAAAGTTTTTGTGAAGATCTGACAGAGGGGAAAAGTTCTCTATTCTCTACTATTC 739
 QY 1019 ACGCCCTT 1026
 DB 740 ATGCTATT 747

RESULT 13
 AAZ60890
 ID AAZ60890 standard; cDNA; 1547 BP.
 XX
 AC AAZ60890;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE cDNA encoding a geranylgeranyl pyrophosphate synthetase (hGGPPS).
 XX
 KW Human; geranylgeranyl pyrophosphate synthetase; hGGPPS; chromosome 1;
 KW 1q42-1q43 locus; prostate cancer; hGGPPS; biallelic marker;
 KW mevalonic biosynthetic pathway; ss.
 XX
 OS Homo sapiens.
 XX

Key Location/Qualifiers
 CDS 218..1120
 FT /*tag= a
 FT /product= "geranylgeranyl pyrophosphate synthetase"
 FT polyA_signal 1422..1427
 FT /*tag= b

WO200005382-A2.
 XX
 XX 03-FEB-2000.
 PD

XX 23-JUL-1999; 99WO-IB001353.
 XX 23-JUL-1998; 98US-0093940P.
 XX (GENT) GENSET.
 XX Bougueleret L;
 XX WPI; 2000-182704/16.
 XX P-PSDB; AAY68909.
 XX New isolated human geranyl-geranyl pyrophosphate synthetase nucleic
 XX acids, used to develop agents for the diagnosis of, e.g. pathologies
 XX related to a defect in the mevalonic biosynthetic pathway.
 XX Claim 3; Page 82-84; 88pp; English.
 XX The present sequence encodes a human geranylgeranyl pyrophosphate
 XX synthetase (hGGPPS). Two differently spliced mRNAs exist for this gene.
 XX The first spliced mRNA is derived from a cDNA (AAZ60888) which comprises
 XX exons 1, 2, 3 and 4. The second mRNA is derived from a cDNA (AAZ60889)
 XX which comprises 1b1s, 2, 3, and 4. The hGGPPS gene is located on
 XX chromosome 1, at the 1q42-1q43 locus. This chromosome 1 locus has been
 XX shown to carry a predisposing gene for prostate cancer. The nucleic acids
 XX encoding hGGPPS can be used for screening for agents which modulate the
 XX expression of the hGGPPS gene. Such agents can be used in therapeutic
 XX applications. The allelic markers associated with the hGGPPS gene can
 XX be used for the diagnosis of diseases related to an alteration in the
 XX regulatory or coding regions of hGGPPS, such as pathologies related to a
 XX defect in the mevalonic biosynthetic pathway. The products can also be
 XX used for detection, diagnosis and drug screening
 XX SQ Sequence 1547 BP; 484 A; 296 C; 315 G; 447 T; 0 U; 5 Other;
 XX Query Match 7.3%; Score 114; DB 3; Length 1547;
 XX Best Local Similarity 52.1%; Pred. No. 7.2e-17;
 XX Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
 QY 359 AAAGCTTGATTTCAAACTTATATACATCTTTTGAACCTGGCAGACTTTAGAC 418
 DB 243 AAAGAAATCTTCTAGAACCTTATATATCTTCTAGTTACCAAGTAAACAAGTGAGAA 302
 QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCAAAGACAGCTGGCCATAG 478
 DB 303 CCAAACTTTCAGGCAATTAATCATCTGGCTGAAAGTTCCAGAGCAAGCTACAGATTA 362
 QY 479 TTTCGCAAAATTTGAGCTTTTGCAATATCCAGCCTTTTATATGAGAGATAGAGATA 538
 DB 363 TTATTAAGAGTACAGAAATGTTGCATAATGCCAGTTTACTCATCATGATGATATTGAAGACA 422
 QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCATTATCTTCGCTGTACCTCCACTA 598
 DB 423 ACTCAAACTCCGAGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 482
 QY 599 TAAACACCGCAATATATATATTTTACAGAGCCTCACTTGTATGCGAGTCAACACAA 658
 DB 483 TCAATCTGCAATTAACGTGATTTTCTTGGCTGGAGAACTCT----- 527
 QY 659 AAGAGCCTTTGATCATATTTGATGATGATTTTCAAGAGAAATGATCAATCAATACATA 718
 DB 528 TAAACCTTTGATCACCAGATGAGTGAAGCTTTTACCAGGCTTTTGGAACTCCATC 587
 QY 719 GGGGACAGGCTTGATATATATCTGAGAGACTTTCTGCTCAAAATCATACCTACTCAGG 778
 DB 588 AGGACAGAGCTGATATATTTACTGGAGGA-----TAAATACACTGTCCCATCTGAG 641
 QY 779 AGATGATTTGATATGTTATGATATTAATAAAGAGGGGCTTTTCAAGATTAACCTTGAGAC 838
 DB 642 AAGAATATAAGCTATGTTGCTGCGAGAAACAGGTGGACTGTTTGGATAGCAGTAGGTC 701
 QY 839 TCATGAAGCGTGTCTCTCTCTCTCACACAGGCGCATTCGTTGGTTCTTTCATATATC 898

DB 702 TCATGCGATTTGTTCTCTGATTACAAAGAAATTTAAACCGCTACTT-----AATA 752
 QY 899 TTCCTGGTATATTTATCAGATTACAGATGATATCTTGAATTTGAAGATTTCCAAATCT 958
 DB 753 CACTTGGGCTCTTTTTCCAAATTAGGATGATTAATGCTAATCTACCTCCAAAGATATA 812
 QY 959 CCAGCGAAAGAGCGCTTGTCTGAGGACATTTACAGAGGGGAAAGTTATCTTTTCCCATCTGCC 1018
 DB 813 GTGAAAACAAAGATTTTGTGAGATCTCAGAGAGGAAAGTTCATCTTCTTCTACTATTC 872
 QY 1019 ACGCCCTT 1026
 DB 873 ATGCTATT 880
 RESULT 14
 ID ADL63550 standard; DNA; 1692 BP.
 AC ADL63550;
 XX 20-MAY-2004 (first entry)
 DT Human ovarian cancer DNA marker #21762.
 XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
 OS Homo sapiens.
 XX WO200170979-A2.
 PD 27-SEP-2001.
 XX 21-MAR-2001; 2001WO-US009126.
 XX 21-MAR-2000; 2000US-0191031P.
 PR 25-MAY-2000; 2000US-0207124P.
 PR 15-JUN-2000; 2000US-0211940P.
 PR 07-JUL-2000; 2000US-0216820P.
 PR 25-JUL-2000; 2000US-0220661P.
 XX 21-DEC-2000; 2000US-0257672P.
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Lee J, Lillie J;
 WPI; 2001-611502/70.
 Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
 cancer cells as compared to their normal non-cancerous ovarian cells are
 used to characterize stage, grade, histological type of ovarian cancer.
 Disclosure; SEQ ID NO 21762; 106pp; English.
 The invention relates to nucleic acid markers which are overexpressed in
 ovarian cancer cells as compared to their expression in normal (i.e. non-
 cancerous) ovarian cells. The invention also relates to polypeptides
 encoded by the markers, antibodies that selectively bind to the
 polypeptides, a method of inhibiting ovarian cancer in a patient at risk
 of developing ovarian cancer involving inhibiting expression of a gene
 corresponding to a marker of the invention and a method of treating a
 patient afflicted with ovarian cancer comprising providing to cells of
 the patient an antisense oligonucleotide complementary to a marker of the
 invention. The markers are useful for assessing if a patient is afflicted
 with ovarian cancer, which involves comparing the level of expression of
 a marker in a patient sample and a normal level of expression of the
 marker in a control non-ovarian cancer sample. A difference between the
 expression levels indicates ovarian cancer. The level of expression of a
 marker corresponds to a secreted protein or to a transcribed
 polynucleotide or its portion. The level of expression of the marker is
 assessed by detecting the presence in the sample, a protein or protein
 fragment corresponding to the marker. The presence of protein or protein
 fragment is detected using an antibody that specifically binds with the

Db 378 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATATTGAAGACA 437
 QY 539 ATGCTCCCTTGTAGAGGGGACAGACCACCTTCTCACCTTATCTTCGGTGTACCCCTCCACATA 598
 Db 438 ACTGAAACTCCGACGTGGCTTCCAGTGGCCACAGCATCTATGGAATCCCACATCTGTCA 497
 QY 599 TAAACACCCGAAATTTATATGTATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAA 658
 Db 498 TCAATTCGCCAATTAACGTGTATTTCTTGGCTTGGAGAAAGTCT-----542
 QY 659 AAGAGCCTTTGTATCAATAATTTGATTAACGATTTTCAAOGAAGAAATGATCAATCTACATA 718
 Db 543 TAAACCTTTGATCACCCAGATGAGTGAAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 602
 QY 719 GGGGACAGGCTTTGGATATATATCTGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778
 Db 603 AGGGACAGGCTTAGAATTTTACTGGAGGA-----7AATTACACTTGTCCACCTGAAG 656
 QY 779 AGATGTATTTGAATATGTTTATGAATAAACAAGGGGCGCTTTTTCAGATTAAACGTTGAGAC 838
 Db 657 AAGAATATAAAGCTATGGTGTGCGAGAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC 716
 QY 839 TCATGGAAGCGTGTCTCTCTCTCTCACACACAGGCCATTCGTTGGTTCCTTTTCATAAATC 898
 Db 717 TCATGCAAGTGTCTCTCTGATTACAAAGAAAGATTTAAACCGCTACTT-----AATA 767
 QY 899 TTCTGGGTATTATTATATCAGATTAGAGATGATTACTTTGAATTTGAAAGATTTCCAAATGT 958
 Db 768 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTTAATCTACATCCAAAGATATA 827
 QY 959 CCAGCGAAAGGCTTTGCTGAGACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCC 1018
 Db 828 GTGAAAACAAAAGTTTTCGTGAAGATCTGACAGAGGGAAGTTCTCAITTTCTCTACTATC 887
 QY 1019 ACGCCCTT 1026
 Db 888 ATGCTATT 895

Search completed: October 16, 2004, 23:38:17
 Job time : 738.231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 18, 2004, 07:18:36 ; Search time 0.001 Seconds
(without alignments)
1051.230 Million cell updates/sec

Title: us-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINDPVWSSQNE.....DTATNLHDELLYIIDHLSL 335

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1 seqs, 1569 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=soft -Ous-10-041-018-22 -DB=us-10-041-018-1
-SUFFIX=pto -OUT=align22_1 -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits
-START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1 -DOCALIGN=200
-THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc
-NRM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO_XLPXY
-NEG SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : us-10-041-018-1.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-----------------|
| 1 | 1720 | 100.0 | 1569 | 1 | us-10-041-018-1 |

ALIGNMENTS

RESULT 1

us-10-041-018-1

Alignment Scores:

| Pred. No.: | Score: | Length: |
|------------------------|---------|-----------------|
| 0 | 1720.00 | 1569 |
| Percent Similarity: | 100.00% | Matches: 335 |
| Best Local Similarity: | 100.00% | Conservative: 0 |
| Query Match: | 100.00% | Mismatches: 0 |
| DB: | 1 | Indels: 0 |
| | | Gaps: 0 |

us-10-041-018-22 (1-335) x us-10-041-018-1 (1-1569)

| Qy | 1 | MetGluAlaLysIleAspGluLeuIleAsnAspProValTrpSerSerGlnAsnGlu 20 |
|----|-----|--|
| Db | 301 | ATGAGGCGCAAGATAGATGAGTCAATATGATCTCTTTGGTCCGAGCAAAATGAA 360 |

| Qy | 21 | SerLeuIleSerLysProTyrAsnHisIleLeuLeuLysProGlyLysAsnPheArgLeu 40 |
|----|------|--|
| Db | 361 | AGCTTGATTTCAAAACCTTATAATCATCTCTTTGAAACCTGGCAAGAACTTTAGACTA 420 |
| Qy | 41 | AsnLeuIleValGlnIleAsnArgValMetAsnLeuProLysAspGlnLeuAlaIleVal 60 |
| Db | 421 | AATTTAATAGTTCAAATTAACAGATTATGAATTTGCCAAAGACCAGCTGGCATAGTT 480 |
| Qy | 61 | SerGlnIleValGluLeuLeuHisAsnSerSerLeuLeuIleAspAspIleGluAspAsn 80 |
| Db | 481 | TGCAAAATTTGAGCTCTTGCAATAATTCAGCTTTTAATCGACGATATAGAAGATAAT 540 |
| Qy | 81 | AlaProLeuArgArgGlyGlnThrSerHisLeuIlePheGlyValProSerThrIle 100 |
| Db | 541 | GCTCCCTTGAGAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600 |
| Qy | 101 | AsnThrAlaAsnTyrMetTyrPheArgAlaMetGlnLeuValSerGlnIleThrLys 120 |
| Db | 601 | ACACCGCAATATATATGATTATTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA 660 |
| Qy | 121 | GluProLeuTyrHisAsnLeuIleThrIlePheAsnGluGluLeuIleAsnLeuHisArg 140 |
| Db | 661 | GAGCTTTTGATCATATAATTTGATTAGATTTCACGAGAAGATTGATCAATCTACATAGG 720 |
| Qy | 141 | GlyGlnGlyIleuAspIleTyrTrpArgAspPheLeuProGluIleIleProThrGlnGlu 160 |
| Db | 721 | GGACAAGGCTTGATATATATCTGAGAGACTTTCTGCCTGAAATCATACTACTCAGGAG 780 |
| Qy | 161 | MetTyrLeuAsnMetValMetAsnLysThrGlyGlyLeuPheArgLeuThrLeuArgLeu 180 |
| Db | 781 | ATGTAATTTGATGTTATGATTAATAAAGACGGCGCTTTTTCAGATTAACTGTAGACTC 840 |
| Qy | 181 | MetGluAlaLeuSerProSerSerHisGlyHisSerLeuValProPheIleAsnLeu 200 |
| Db | 841 | ATGGAAGCGCTCTCTCTCTCTCACACACGCGCAATTCGTTCGTTCTCTTCAATAATCTT 900 |
| Qy | 201 | LeuGlyIleIleTyrGlnIleArgAspAspTyrLeuAsnLeuLysAspPheGlnMetSer 220 |
| Db | 901 | CTGGGTATTTATTTATCAGATTAGATGATTAATTTGAAATTTGAAAGATTTCCAAATGTC 960 |
| Qy | 221 | SerGluLysGlyPheAlaGluAspIleThrGluGlyLysLeuSerPheProIleValHis 240 |
| Db | 961 | ACGAAAAAGGCTTTGCTGAGACATTACAGAGGGAAGTTATCTTTTCCCATCGTCCAC 1020 |
| Qy | 241 | AlaLeuAsnPheThrLysThrLysGlyGlnThrGluGlnHisAsnGluIleLeuArgIle 260 |
| Db | 1021 | GCCCTTAACCTTCACTTAAACGAAAGGTCAAACTGAGCAACACATGAATTTCTAAGAAAT 1080 |
| Qy | 261 | LeuLeuLeuArgThrSerAspLysAspIleLysLeuLysLeuIleGlnIleLeuGluPhe 280 |
| Db | 1081 | CTCCTGTTGAGGACAAGTGATAAAAGATATAAACTTAAAGCTGATTCAAATACTGGAATTC 1140 |
| Qy | 281 | AspThrAsnSerLeuAlaTyrThrLysAsnPheIleAsnGlnLeuValAsnMetIleLys 300 |
| Db | 1141 | GACACCAATTCATTGGCCCTACACCAAAATTTTATTAATCAATTAGTGAATATGATAAAA 1200 |
| Qy | 301 | AsnAspAsnGluAsnLysTyrLeuProAspLeuAlaSerHisSerAspThrAlaThrAsn 320 |
| Db | 1201 | AATGATAATGAAATAAGTATTTTACCTGATTTGGCTTCGCATTCGACACCGCCCAACAT 1260 |
| Qy | 321 | LeuHisAspGluLeuLeuTyrIleIleAspHisLeuSerGluLeu 335 |
| Db | 1261 | TTACATGACGAATTTGTTATATATATATATAGACCACTTATCCGAATTG 1305 |

Search completed: October 18, 2004, 07:18:37

Job time : 1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 18, 2004, 07:20:37 ; Search time 2 Seconds
(without alignments)
2.189 Million cell updates/sec

Title: us-10-041-018-383

Perfect score: 4113

Sequence: 1 MNLSLCIASPLLTGKSRPAA.....TVKDIINYNPVLVINEBQR 784

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1 seqs, 2792 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=soft -Q=us-10-041-018-383 -DB=us-10-041-018-361
-SUFFIX=ptc -OUT=align383_361 -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0 -UNITS=bits
-START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi -LIST=1 -DOCALIGN=200
-THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO_XLPXY
-NEG SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : us-10-041-018-361.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length DB | ID | Description |
|------------|-------------|-----------|------|---------------------|
| 1 | 4113 | 100.0 | 2792 | 1 us-10-041-018-361 |

ALIGNMENTS

RESULT 1

us-10-041-018-361

Alignment Scores:

| Pred. No.: | Score: | Length: | Matches: |
|------------------------|---------|---------------|----------|
| 0 | 4113.00 | 2792 | 784 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DB: | 1 | Gaps: | 0 |

us-10-041-018-383 (1-784) x us-10-041-018-361 (1-2792)

Qy 1 MetAsnLeuSerLeuCysIleAlaSerProLeuLeuThrLysSerAsnArgProAlaAla 20

Db 174 ATGAATCTTTCACATATGCGTCCCTTTGTTAAACCAATCAAATCGACCGCGGT 233

Qy 21 LeuSerAlaIleHisThrAlaSerThrSerHisGlyGlyGlnThrAsnProThrAsnLeu 40
Db 234 CTGTGAGCTATTTCATACAGCATCAACTTCACATGGTGGACAACTAATCCCACTAATCTG 293

Qy 41 IleIleAspThrThrIlyeGluArgIleGlnIlyeGlnPheLysAsnValGluIleSerVal 60
Db 294 ATCATTGATACAAACCAAGACGGATCCAAACACAGTTTAAAAATGTAGAAATTTCTGTT 353

Qy 61 SerSerTyrAspThrAlaTrpValAlaMetValProSerProAsnSerProLysSerPro 80
Db 354 TCTTCATATGACACAGCATGGTAGCCATGGTCCCTTCTCCAACTCACCACAAATGCCCT 413

Qy 81 CysPheProGluCysLeuAsnTrpLeuIleAsnAsnGlnLeuAsnAspGlySerTrpGly 100
Db 414 TGTTCCTCCCTGAGTGTCTCAATTTGGTTAATAATAATCAGCTTAATGATGGTTTCATGGGT 473

Qy 101 LeuValAsnHisThrHisAsnHisProLeuLeuLysAspSerLeuSerSerThr 120
Db 474 CTGTGTAATCACACTCAATAATCACCGGTTGCTTAAAGATTCTCTATCTTCAACA 533

Qy 121 LeuAlaCysIleValAlaLeuLysArgTrpAsnValGlyGluAspGlnIleAsnLysGly 140
Db 534 TTAGCATGTAATTTGTCATTAAGAGATGGATGTTGGGAGACATCAATAAATAAGT 593

Qy 141 LeuSerPheIleGluSerAsnLeuAlaSerAlaThrGluLysSerGlnProSerProIle 160
Db 594 CTAAGTTTTATTGAGTCAATCTTGTCTCAGCTACTGAAAAAAGTCAACCATCTCCCAT 653

Qy 161 GlyPheAspIleIlePheProGlyLeuLeuGluTyrAlaLysAsnLeuAspIleAsnLeu 180
Db 654 GGTTCATCATCAATTTCTCGTTTGTGAGTATGCGAAAACTTGACATAACCTC 713

Qy 181 LeuSerLysGlnThrAspPheSerLeuMetLeuHisLysArgGluLeuGluGlnLysArg 200
Db 714 CTTTCAAAACAAACAGATTTTAGTTGATGTACATAAGAGGAATTCGAGCAAAAAAGA 773

Qy 201 CysHisSerAsnGluMetAspGlyTyrLeuAlaTyrIleSerGluGlyLeuGlyAsnLeu 220
Db 774 TGCAATTCAAATGAGATGGATGGATCTTGGCGTATATCTCTGAAGGACTCGGTAATTTA 833

Qy 221 TyrAspTrpAsnMetValLysLysTyrGlnMetLysAsnGlySerValPheAsnSerPro 240
Db 834 TATGATTTGGAATATGGTGAAGAAATATCAGATGAAAAATGGTTCTGTTTCACTCACA 893

Qy 241 SerAlaThrAlaAlaAlaPheIleAsnHisGlnAsnProGlyCysLeuAsnTyrLeuAsn 260
Db 894 TCAGCAACAGCTGCTGCTTCATTAATCATCAAAATCTCGTTGTTCTTAATTAATTAAT 953

Qy 261 SerLeuLeuAspLysPheGlyAsnAlaValProThrValTyrProHisAspLeuPheIle 280
Db 954 TCACCTTTGGCAAGTTTGGTAATGCGAGTCCCAACAGTTTATCTCATGATTTATTTATC 1013

Qy 281 ArgLeuSerMetValAsnThrIleGluArgLeuGlyIleSerHisHisPheArgValGlu 300
Db 1014 CGACTTTCTATGGTTGACAAATTTGAAGATTTAGGAATTTCCACCATTTCAGAGTGGAA 1073

Qy 301 IleLysAsnValLeuAspGluThrTyrArgCysTrpValGluArgAspGluGlnIlePhe 320
Db 1074 ATTAATAAATGTTTATAGATGAACACATACAGATGTTGGTGGACAGAGATGACAAATATTC 1133

Qy 321 MetAspValValThrCysAlaLeuAlaPheArgLeuLeuArgIleAsnGlyTyrGluVal 340
Db 1134 ATGGATGTTGTAAACATGCTTTAGCCCTTTCCGTTTATTAAGGATCAATGGGTATGAAGTT 1193

Qy 341 SerProAspProLeuAlaGluIleThrAsnGluLeuAlaLeuLysAspGluTyrAlaAla 360
Db 1194 TCCCAGATCCATGGCTGAAATTTACTAATGAATAGCTTTGAAAAGACGAATATGCGAGCT 1253

Qy 361 LeuGluThrTyrHisAlaSerHisIleLeuTyrGlnGluAspLeuSerSerGlyLysGln 380
Db 1254 CTTGAAACATATCATGCGTCACATATATTATATCAACAGAGGATTTATCTTCTGGAAAAACA 1313

```
QY 381 lleLeuIysSerAlaaspPheLeuIysGluilleSerThrAspSerAsnArgLeuSer 400
Db 1314 ATCTTGAAGTCAGCTGATTCTCTCAAGAGATTAATATCCCTGATTCAAACAGGCTTCT 1373
QY 401 LysLeuIleHisIysGluValGluAsnAlaLeuLysPheProIleAsnThrGlyLeuGlu 420
Db 1374 AAATTAATTCACAAAGAGGTGGAAATGCTCTTAAGTTCCTTCAATCAATACCGGTTAGAA 1433
QY 421 ArgIleAsnThrArgArgAsnIleGlnLeuTyrAsnValAspAsnThrArgIleLeuLys 440
Db 1434 CGCATAAACACTAGACGAATATACAGCTTCAATGTAGACAAATCAAGAAATTCGAAA 1493
QY 441 ThrThrTyrHisSerSerAsnIleSerAsnThrAspTyrLeuArgLeuAlaValGluAsp 460
Db 1494 ACTACATATACATCATCAATATTAGTAACACTGATTAACCTAAGCTTGGCTGTGAAGAT 1553
QY 461 PheTyrThrCysGlnSerIleTyrArgGluGluLeuLysGlyLeuGluArgTrpValVal 480
Db 1554 TTCTACACCTGCCAATCTATTATTATCGTGAAGAAATTAAGAGCTCTTGAAAGGTGGGTGTA 1613
QY 481 GluAsnLysLeuAspGlnLeuIysPheAlaArgGlnIysThrAlaTyrCysTyrPheSer 500
Db 1614 GAGATAAAGTTGACACAGCTCAAGTTTGCTAGCAAAAGACCGCCTACTGTATTCTCT 1673
QY 501 ValAlaAlaThrLeuSerSerProGluLeuSerAspAlaArgIleSerTrpAlaLysAsn 520
Db 1674 GTTGCTGCAACACTTTCGTCCTCCGAATATACAGTCCGGTATTTCAITGGGCCAAAAAT 1733
QY 521 GlyIleLeuThrThrValValAspAspPhePheAspIleGlyGlyThrIleAspGluLeu 540
Db 1734 GGCATATTAACTACAGTAGTTGATGACTTTTGTGATATCGGTGGTACAATCGATGAATTG 1793
QY 541 ThrAsnLeuIleGlnCysValGluLysTrpAsnValAspValAspLysAspCysSer 560
Db 1794 ACCAACCCTGATTCATGTGTGTAATAATGCAATGTAGATGTCGACAGGATTTGTGTCA 1853
QY 561 GluHisValArgIleLeuPheLeuAlaLeuLysAspAlaIleCysTrpIleGlyAspGlu 580
Db 1854 GAGCATGTTCCGATTTTATTTTATTTAGCATTAAAGATGCAATCTGTGGATTGGAGATGAA 1913
QY 581 AlaPheLysTrpGlnAlaArgAspValThrSerHisValIleGlnThrTrpLeuGluLeu 600
Db 1914 GCCTTTAAATGGCAAGCGCGCATGTAACTAGCCATGTTATTCAAACTTGGTTGGAACATA 1973
QY 601 MetAsnSerMetLeuArgGluAlaIleTrpThrArgAspAlaTyrValProThrLeuAsn 620
Db 1974 ATGAATAGTATGTGAGAGAAGCTATATGACAGAGATGCTTATGTGCCAACATTAAT 2033
QY 621 GluTyrMetGluAsnAlaTyrValSerPheAlaLeuGlyProIleValLysProAlaIle 640
Db 2034 GAATATATGGAACCGCTTACGTGTCATTTGCTAGCCCGATTTGCAAGCCGGCTATT 2093
QY 641 TyrPheValGlyProLysLeuSerGluGluIleValGlnSerSerGluTyrHisAsnLeu 660
Db 2094 TACTTTGTGGGGCCCAATATATCAGAGAGATTTGTGAAAAGCTCTGAATATCATTAATCTA 2153
QY 661 PheLysLeuMetSerThrGlnGlyArgLeuLeuAsnAspIleHisSerPheLysArgGlu 680
Db 2154 TTTAAGCTAATAGACACGCGGGTCGACTTCTTAACGATATCCATAGCTTCAAGAGGAA 2213
QY 681 PheLysGluGlyLysLeuAsnAlaValAlaLeuHisLeuSerAsnGlyGluSerGlyLys 700
Db 2214 TTTAAGGAGGCAAAATTAACCGCGGTAGCATTTGATTTGAGTAACGGAGAAAGTGGGAAA 2273
QY 701 ValGluGluValValGluGluMetMetMetIleLysAsnLysArgLysGluLeu 720
Db 2274 GTGGAAGAAGAGGTTGTGCGAGCATGATGATGATGATTAATAACAGAGAGAAAGAAATTA 2333
QY 721 MetLysLeuIlePheGluGluAsnGlySerIleValProArgAlaCysLysAspAlaPhe 740
Db 2334 ATGAATTAATTTTGAAGAAAATGGTAGCATTTGTTCTAGAGCTTGTAAAGATGCATTT 2393
QY 741 TrpAsnMetCysHisValLeuAsnPhePheTyrAlaAsnAspGlyPheThrGlyAsn 760
```

```
Db 2394 TCGAACATGTGTCAAGTGTGTAATTTTTTTTACGCAAAACGATGACGGGTTTACTGAAAC 2453
QY 761 ThrIleLeuAspThrValLysAspIleIleTyrAsnProLeuValLeuValAsnGluAsn 780
Db 2454 ACGATTCCTTGATACCTGTGAGGACATCATTTACACCCCGTTGGTGTGTGTAATGAAAT 2513
QY 781 GluGluGlnArg 784
Db 2514 GAAGAACAAAGG 2525
```

Search completed: October 18, 2004, 07:20:45
Job time : 8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 15:57:41 ; Search time 6603.4 Seconds
(without alignments)
11236.263 Million cell updates/sec

Title: US-10-041-018-1
Perfect score: 1569
Sequence: 1 aatattacatatagatag.....aatgcggaattgcga 1569

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hlg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_to.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|--------|---------------|--------|----|-------------|--------------------|
| 1 | 1569 | 100.0 | 1569 | 6 | AR071897 | AR071897 Sequence |
| 2 | 1569 | 100.0 | 1569 | 8 | SCU31632 | U31632 Saccharomyc |
| 3 | 1356.4 | 86.4 | 37947 | 8 | SCU39205 | U39205 Saccharomyc |
| 4 | 1008 | 64.2 | 1008 | 6 | BD177419 | BD177419 A method |
| 5 | 1008 | 64.2 | 1008 | 6 | BD170961 | BD170961 Process f |
| 6 | 1008 | 64.2 | 1008 | 6 | BD171079 | BD171079 Process f |
| 7 | 1008 | 64.2 | 1008 | 8 | AY692852 | AY692852 Saccharom |
| 8 | 1005 | 64.1 | 1005 | 6 | AR071896 | AR071896 Sequence |
| 9 | 277.2 | 17.7 | 110000 | 8 | CR380954_04 | Continuation (5 of |
| 10 | 277.2 | 17.7 | 110000 | 8 | CR380954_05 | Continuation (6 of |
| 11 | 261 | 16.6 | 35868 | 8 | SCU41849 | U41849 Saccharomyc |
| 12 | 183.4 | 11.7 | 300316 | 8 | AR016894 | AR016894 Eremothec |
| 13 | 128.6 | 8.2 | 952 | 11 | CNS06EHJ | AL395117 T3 end of |
| 14 | 121.6 | 7.8 | 1130 | 10 | AB118242 | AB118242 Rattus no |
| 15 | 121.6 | 7.8 | 1487 | 10 | AB118243 | AB118243 Rattus no |
| 16 | 121.6 | 7.8 | 1686 | 10 | AB118241 | AB118241 Rattus no |
| 17 | 121.6 | 7.8 | 2616 | 10 | AB118240 | AB118240 Rattus no |
| 18 | 121.6 | 7.8 | 2950 | 10 | AB118239 | AB118239 Rattus no |
| 19 | 121.6 | 7.8 | 3172 | 10 | AB118238 | AB118238 Rattus no |

| | | | | | | |
|----|-------|-----|--------|----|-------------|--------------------|
| 20 | 120.2 | 7.7 | 110000 | 8 | CR382125_20 | Continuation (21 o |
| 21 | 117 | 7.5 | 110000 | 8 | CR382130_20 | Continuation (21 o |
| 22 | 115.6 | 7.4 | 1411 | 9 | BC067768 | BC067768 Homo sapi |
| 23 | 114 | 7.3 | 903 | 6 | BD178210 | BD178210 Human ger |
| 24 | 114 | 7.3 | 903 | 9 | AB017971 | AB017971 Homo sapi |
| 25 | 114 | 7.3 | 1256 | 6 | CQ17435 | CQ17435 Sequence |
| 26 | 114 | 7.3 | 1387 | 9 | AF125394 | AF125394 Homo sapi |
| 27 | 114 | 7.3 | 1395 | 9 | AB016043 | AB016043 Homo sapi |
| 28 | 114 | 7.3 | 1466 | 9 | AF057698 | AF057698 Homo sapi |
| 29 | 114 | 7.3 | 1489 | 9 | BC005252 | BC005252 Homo sapi |
| 30 | 114 | 7.3 | 1692 | 6 | CQ414691 | CQ414691 Sequence |
| 31 | 114 | 7.3 | 2534 | 9 | AK025139 | AK025139 Homo sapi |
| 32 | 114 | 7.2 | 1025 | 10 | AB016044 | AB016044 Mus muscu |
| 33 | 113.6 | 7.2 | 2387 | 10 | BC069913 | BC069913 Mus muscu |
| 34 | 113.6 | 7.2 | 2387 | 10 | BC069913 | BC069913 Mus muscu |
| 35 | 112.4 | 7.2 | 1437 | 9 | AB019036 | AB019036 Homo sapi |
| 36 | 110.4 | 7.0 | 2268 | 10 | BC006798 | BC006798 Mus muscu |
| 37 | 105.8 | 6.7 | 2225 | 8 | GFGPPSGN | X96943 G.fujikuroi |
| 38 | 104.4 | 6.7 | 163216 | 2 | AC122557 | AC122557 Mus muscu |
| 39 | 104.4 | 6.7 | 339779 | 2 | AC121558 | AC121558 Mus muscu |
| 40 | 102.6 | 6.5 | 2511 | 8 | AF479566 | AF479566 Aspergill |
| 41 | 99.6 | 6.3 | 1032 | 6 | BD189547 | BD189547 Aphidicol |
| 42 | 99.6 | 6.3 | 1332 | 8 | AB079897 | AB079897 Phoma bet |
| 43 | 99.6 | 6.3 | 1888 | 6 | CQ493258 | CQ493258 Sequence |
| 44 | 99.6 | 6.3 | 15629 | 8 | AB114137 | AB114137 Phoma bet |
| 45 | 93.6 | 6.0 | 13601 | 10 | AB118237 | AB118237 Rattus no |

ALIGNMENTS

RESULT 1
AR071897
LOCUS AR071897 1569 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 3 from patent US 5912154.
ACCESSION AR071897
VERSION AR071897.1 GI:7222785
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1569)
AUTHORS Ferro-Novick,S. and Jiang,Y.
TITLE Geranylgeranyl diphosphate synthase proteins, nucleic acid molecules and uses thereof
JOURNAL Patent: US 5912154-A 3 15-JUN-1999;
FEATURES Location/Qualifiers
source 1. 1569
/organism="unknown"
/mol_type="unassigned DNA"

| | | | | |
|-----------------------|-----------------|---|-----------|--------------|
| Query Match | 100.0%; | Score 1569; | DB 6; | Length 1569; |
| Best Local Similarity | 100.0%; | Pred. No. 3.7e-309; | Indels 0; | Gaps 0; |
| Matches 1569; | Conservative 0; | Mismatches 0; | | |
| Qy | 1 | AATATTACATATAGATATAGGACCAAGCCGCAATTTTCATCTGAAAGTAACTCTATT | 60 | |
| Db | 1 | AATATTACATATAGATATAGGACCAAGCCGCAATTTTCATCTGAAAGTAACTCTATT | 60 | |
| Qy | 61 | ATTATAGTGGTATCCACGTTTCACCGTTTCAGCATGACGAAATTTAGCTGTTTTCAT | 120 | |
| Db | 61 | ATTATAGTGGTATCCACGTTTCACCGTTTCAGCATGACGAAATTTAGCTGTTTTCAT | 120 | |
| Qy | 121 | ATGTTATGCTGATCATTTGATGCTTACTACCATTTTCTTGTCTCGCTTCCTTCTTT | 180 | |
| Db | 121 | ATGTTATGCTGATCATTTGATGCTTACTACCATTTTCTTGTCTCGCTTCCTTCTTT | 180 | |
| Qy | 181 | GACGTTTTTTTGAAGCAAAAAAAGTCAAGACAGATGCTGTTTACAAAACCATGTAAGGC | 240 | |
| Db | 181 | GACGTTTTTTTGAAGCAAAAAAAGTCAAGACAGATGCTGTTTACAAAACCATGTAAGGC | 240 | |
| Qy | 241 | TCATTTCAAAGAGCTACTAATAAGAGACAAAGAGTTTACGAGTCTGGAATACTCA | 300 | |

| | | | |
|----|------|---|------|
| Qy | 61 | ATTATAGTGGTATCCAAAGTTTCCAGCGTTCAGCATACGAGAAATTTAGTGTCTTTTGGCAT | 120 |
| Db | 61 | | |
| Qy | 61 | ATTATAGTGGTATCCAAAGTTTCCAGCGTTCAGCATACGAGAAATTTAGTGTCTTTTGGCAT | 120 |
| Db | 61 | | |
| Qy | 121 | ATGTTATGCTGATCATTTGTATGCTTACTACCAATTTTCTTTGCTTCGCCTTGCCTTCCTTT | 180 |
| Db | 121 | | |
| Qy | 121 | ATGTTATGCTGATCATTTGTATGCTTACTACCAATTTTCTTTGCTTCGCCTTGCCTTCCTTT | 180 |
| Db | 121 | | |
| Qy | 181 | GACGTTTTTTTGAAGCAAAAAAAGTCAAGACAGATGCTTTACAAAACCATGTAAGGC | 240 |
| Db | 181 | | |
| Qy | 181 | GACGTTTTTTTGAAGCAAAAAAAGTCAAGACAGATGCTTTACAAAACCATGTAAGGC | 240 |
| Db | 181 | | |
| Qy | 241 | TCATTTTCAAGAAGCTACTAATAGAAGAGAACTTACGAGTTTACGAGTCTTGGAATAATCA | 300 |
| Db | 241 | | |
| Qy | 241 | TCATTTTCAAGAAGCTACTAATAGAAGAGAACTTACGAGTTTACGAGTCTTGGAATAATCA | 300 |
| Db | 241 | | |
| Qy | 301 | ATGAGGCCCAAGATAGATGAGCTGATCAATAATGATCCTGTTTGGTCCAGCGCAAAATCAA | 360 |
| Db | 301 | | |
| Qy | 301 | ATGAGGCCCAAGATAGATGAGCTGATCAATAATGATCCTGTTTGGTCCAGCGCAAAATCAA | 360 |
| Db | 301 | | |
| Qy | 361 | AGCTTGATTTCAAAACCTTATAATCAATCCTTTTGAACCTGGCAAGAACTTTAGACTA | 420 |
| Db | 361 | | |
| Qy | 361 | AGCTTGATTTCAAAACCTTATAATCAATCCTTTTGAACCTGGCAAGAACTTTAGACTA | 420 |
| Db | 361 | | |
| Qy | 421 | AATTTAATAGTTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT | 480 |
| Db | 421 | | |
| Qy | 421 | AATTTAATAGTTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT | 480 |
| Db | 421 | | |
| Qy | 481 | TCGCAAAATTTGTGAGCTCTTCATATAATTCAGACCTTTTAAATCGACGATATAGAAGATAAT | 540 |
| Db | 481 | | |
| Qy | 481 | TCGCAAAATTTGTGAGCTCTTCATATAATTCAGACCTTTTAAATCGACGATATAGAAGATAAT | 540 |
| Db | 481 | | |
| Qy | 541 | GCTCCCTTGAGAAGGGGACAGACACTTCTCACTAATCTTCGGTGATACCTCCACTATA | 600 |
| Db | 541 | | |
| Qy | 541 | GCTCCCTTGAGAAGGGGACAGACACTTCTCACTAATCTTCGGTGATACCTCCACTATA | 600 |
| Db | 541 | | |
| Qy | 601 | AACACCCCAAAATATATGTAATTTAGTACGAGCCATGCAACTTGATTCGAGCTAACCCAAAA | 660 |
| Db | 601 | | |
| Qy | 601 | AACACCCCAAAATATATGTAATTTAGTACGAGCCATGCAACTTGATTCGAGCTAACCCAAAA | 660 |
| Db | 601 | | |
| Qy | 661 | GAGCCTTTGATCATATAATTTGATTACGAATTTTCAACGAAGAATTTGATCAATCTACATAGG | 720 |
| Db | 661 | | |
| Qy | 661 | GAGCCTTTGATCATATAATTTGATTACGAATTTTCAACGAAGAATTTGATCAATCTACATAGG | 720 |
| Db | 661 | | |
| Qy | 721 | GGCAAGGCTTGGATATATATCTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGGAG | 780 |
| Db | 721 | | |
| Qy | 721 | GGCAAGGCTTGGATATATATCTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGGAG | 780 |
| Db | 721 | | |
| Qy | 781 | ATGTAATTTGAATATGGTTATGAATAAAACAGGCGGCCTTTTTCAGATTAACTTGAGACTC | 840 |
| Db | 781 | | |
| Qy | 781 | ATGTAATTTGAATATGGTTATGAATAAAACAGGCGGCCTTTTTCAGATTAACTTGAGACTC | 840 |
| Db | 781 | | |
| Qy | 841 | ATGGAAGCGTGTCTCCTTCTCACACACGCGCATTCGTTGGTTCCTTTTCATAAAATCTT | 900 |
| Db | 841 | | |
| Qy | 841 | ATGGAAGCGTGTCTCCTTCTCACACACGCGCATTCGTTGGTTCCTTTTCATAAAATCTT | 900 |
| Db | 841 | | |
| Qy | 901 | CTGGGTATTTATTCAGATTAGAGATGATTTGTAATTTGAAGATTTTCCAAATGTCC | 960 |
| Db | 901 | | |
| Qy | 901 | CTGGGTATTTATTCAGATTAGAGATGATTTGTAATTTGAAGATTTTCCAAATGTCC | 960 |
| Db | 901 | | |
| Qy | 961 | AGCGAAAAAGCTTTGCTGAGGACATTTACAGAGGGAGTTATCTTTTCCCATCGTCCAC | 1020 |
| Db | 961 | | |
| Qy | 961 | AGCGAAAAAGCTTTGCTGAGGACATTTACAGAGGGAGTTATCTTTTCCCATCGTCCAC | 1020 |
| Db | 961 | | |
| Qy | 1021 | GCCCTTAATCTCACTAAAAAGGTTCAAACTGAGCAACAGTGAAGATTTCTAAGAAATT | 1080 |
| Db | 1021 | | |
| Qy | 1021 | GCCCTTAATCTCACTAAAAAGGTTCAAACTGAGCAACAGTGAAGATTTCTAAGAAATT | 1080 |
| Db | 1021 | | |
| Qy | 1081 | CTCCTGTTGAGGCAAGTGAATAAATAAATCAAGCTGATTTCAAACTGGAATTC | 1140 |
| Db | 1081 | | |
| Qy | 1081 | CTCCTGTTGAGGCAAGTGAATAAATAAATCAAGCTGATTTCAAACTGGAATTC | 1140 |
| Db | 1081 | | |

[illegible]

| | | |
|----------------------------|--|-----------------|
| CDS | /gene="YPL062W" 9061..9465 /gene="YPL062W" /codon_start=1 /product="Ypl062wp" /protein_id="AAB68303.1" /db_xref="GI:1079680" /translation="MTELQYVKGEDTIVETATSTPWLMSPLARAAEKRGSLFPDIN EGHQHVDIVAYGVSKPKRRSPHPGKNKAADKKRTEKEPERKRVGVAPRCKKICI OLLRYRNFVLGIGNVCNIDLLGLTYSAINAI" 9754..11256 /gene="ALD6" 9754..11256 /gene="ALD6" /notes="YPL061W" /codon_start=1 | |
| gene | | |
| CDS | | |
| Query Match | 86.4%; Score 1356.4; DB 8; Length 37947; | |
| Best Local Similarity | 99.9%; Pred. No. 4.3e-266; | |
| Matches 1357; Conservative | 0; Mismatches 1; Indels 0; Gaps 0; | |
| QY | 1 AATATTACATATAGATATAGGCAAGCCGCGCATTTTCATACCTCAAGGTAAACTTCTATT | 60 |
| DB | 1358 AATATTACATATAGATATAGGCAAGCCGCGCATTTTCATACCTCAAGGTAAACTTCTATT | 1299 |
| QY | 61 ATTATAGTGGTATCCAAAGCTTCAACCGCTTCCAGCATAGCAGAAATTACGCTGTTTTTGGCAT | 120 |
| DB | 1298 ATTATAGTGGTATCCAAAGCTTCAACCGCTTCCAGCATAGCAGAAATTACGCTGTTTTTGGCAT | 1239 |
| QY | 121 ATGTATGCTGATCATTTGATGCTTACTACCAATTTTCTTGGCTTCGCTTCCTTCTTT | 180 |
| DB | 1238 ATGTATGCTGATCATTTGATGCTTACTACCAATTTTCTTGGCTTCGCTTCCTTCTTT | 1179 |
| QY | 181 GACGTTTTTTTCAAGCAAAAAAAGTCAAGCAGATGCTTACAAAACCATGTAAAGC | 240 |
| DB | 1178 GACGTTTTTTTCAAGCAAAAAAAGTCAAGCAGATGCTTACAAAACCATGTAAAGC | 1119 |
| QY | 241 TCATTTTCAAGAGAGCTACTAATAGAAAAGAGAAAGATTTTACGAGTCTCGAAAAATCA | 300 |
| DB | 1118 TCATTTTCAAGAGAGCTACTAATAGAAAAGAGAAAGATTTTACGAGTCTCGAAAAATCA | 1059 |
| QY | 301 ATGGAGGCAAGATAGATGAGTGTATCAATATGATTCCTGTTGGTCCAGCCCAAAATGAA | 360 |
| DB | 1058 ATGGAGGCAAGATAGATGAGTGTATCAATATGATTCCTGTTGGTCCAGCCCAAAATGAA | 999 |
| QY | 361 AGCTTGATTTCAAAACCTTATATACATCCCTTTTGAACCTGCGAAGCACTTTAGACTA | 420 |
| DB | 998 AGCTTGATTTCAAAACCTTATATACATCCCTTTTGAACCTGCGAAGCACTTTAGACTA | 939 |
| QY | 421 AATTATATAGTTCAAATTAACAGAGTTATGAAATTTGCCCAAGACACGCTGGCCATAGTT | 480 |
| DB | 938 AATTATATAGTTCAAATTAACAGAGTTATGAAATTTGCCCAAGACACGCTGGCCATAGTT | 879 |
| QY | 481 TCGCAATTTGTAGCTCTTGCATTAATTCAGACCTTTTAAATCGACATATAGAGATAAT | 540 |
| DB | 878 TCGCAATTTGTAGCTCTTGCATTAATTCAGACCTTTTAAATCGACATATAGAGATAAT | 819 |
| QY | 541 GTCCTCTTGAGAGGGGACAGACCTTCTACATTAATTTCTGGGTGACCTTCCACTATA | 600 |
| DB | 818 GTCCTCTTGAGAGGGGACAGACCTTCTACATTAATTTCTGGGTGACCTTCCACTATA | 759 |
| QY | 601 AACACCGCAAAATTATGATTTTACAGAGCCATGCAATTTGATTCGACGTAACCAACAAAA | 660 |
| DB | 758 AACACCGCAAAATTATGATTTTACAGAGCCATGCAATTTGATTCGACGTAACCAACAAAA | 699 |
| QY | 661 GAGCCTTTGTATCATATATTTGATTCAGATTTTCAACGAAGATTTGATCAATCTACATAGG | 720 |
| DB | 698 GAGCCTTTGTATCATATATTTGATTCAGATTTTCAACGAAGATTTGATCAATCTACATAGG | 639 |
| QY | 721 GGACAGGCTTGGATATATATCTGGAGACCTTCTGCTCGTAATCATATCTACTCAGGAG | 780 |
| DB | 638 GGACAGGCTTGGATATATATCTGGAGACCTTCTGCTCGTAATCATATCTACTCAGGAG | 579 |
| QY | 781 AUGTATTGAATATGCTTATGATATAAACAAGGGGCGCTTTTCAGATTAACTGAGACTC | 840 |
| DB | 578 AUGTATTGAATATGCTTATGATATAAACAAGGGGCGCTTTTCAGATTAACTGAGACTC | 519 |
| QY | 841 ATGGAAGCGCTGCTCTCTCTCACACCAAGGGCATTTCGTTGCTTCCCTTTCAATAATCTT | 900 |
| DB | 518 ATGGAAGCGCTGCTCTCTCTCACACCAAGGGCATTTCGTTGCTTCCCTTTCAATAATCTT | 459 |
| QY | 901 CTGGGTATTATTATATCAGATTAGAGATGATTCTTGAATTTGAAAGATTTCCAAATGTCC | 960 |
| DB | 458 CTGGGTATTATTATCAGATTAGAGATGATTCTTGAATTTGAAAGATTTCCAAATGTCC | 399 |
| QY | 961 AGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTTCCATCTGTCAC | 1020 |
| DB | 398 AGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTTCCATCTGTCAC | 339 |
| QY | 1021 GCCCTTTAACTTCACTAAAAAGGCTCAAACTGAGCAACACAAATGAAATTTCTAAGAAAT | 1080 |
| DB | 338 GCCCTTTAACTTCACTAAAAAGGCTCAAACTGAGCAACACAAATGAAATTTCTAAGAAAT | 279 |
| QY | 1081 CTCCTGTTGAGGACAGTGAATAAGATATAAACTAAAGCTGATTCAAAATCTGGAATTC | 1140 |
| DB | 278 CTCCTGTTGAGGACAGTGAATAAGATATAAACTAAAGCTGATTCAAAATCTGGAATTC | 219 |
| QY | 1141 GACACCAATTCATTCGCTTACACCAAAATTTTATTAATCAATTAGTGAATATGATAAAA | 1200 |
| DB | 218 GACACCAATTCATTCGCTTACACCAAAATTTTATTAATCAATTAGTGAATATGATAAAA | 159 |
| QY | 1201 AATGATAATGAAAAAAGTATTACCTGATTGCTTCCGATTCGACACCCGCCCAAT | 1260 |
| DB | 158 AATGATAATGAAAAAAGTATTACCTGATTGCTTCCGATTCGACACCCGCCCAAT | 99 |
| QY | 1261 TTACATGACGAATTCCTTATATATAATAGACCACTTATCCGAATTTGAAAAATTAATGATC | 1320 |
| DB | 98 TTACATGACGAATTCCTTATATATAATAGACCACTTATCCGAATTTGAAAAATTAATGATC | 39 |
| QY | 1321 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATAAGGCC | 1358 |
| DB | 38 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATAAGGCC | 1 |
| RESULT 4 | | |
| BD177419 | 1008 bp DNA linear | PAT 16-APR-2003 |
| LOCUS | | |
| DEFINITION | A method of producing prenyl alcohol. | |
| ACCESSION | BD177419.1 | |
| VERSION | GI:30014680 | |
| KEYWORDS | JP 2002300896-A/12. | |
| SOURCE | Saccharomyces cerevisiae (baker's yeast) | |
| ORGANISM | Saccharomyces cerevisiae | |
| REFERENCE | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | |
| AUTHORS | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | |
| TITLE | 1 (bases 1 to 1008) | |
| JOURNAL | Muramatsu, M., Kobata, M. and Shimizu, A. | |
| COMMENT | A method of producing prenyl alcohol | |
| | Patent: JP 2002300896-A 12 15-OCT-2002; | |
| | TOYOTA MOTOR CORP | |
| | OS Saccharomyces cerevisiae (yeast) | |
| | PN JP 2002300896-A/12 | |
| | PD 15-OCT-2002 | |
| | PF 18-JAN-2002 JP 2002010528 | |
| | PI MASAYOSHI MURAMATSU, MITSUO KOBATA, AKIRA SHIMIZU PC | |
| | Cl2P7/04// (Cl2P7/04, Cl2R1:85), (Cl2P7/04, Cl2R1:78), (Cl2P7/04, PC | |
| | Cl2R1:84), | |
| | PC (Cl2P7/04, Cl2R1:72), (Cl2P7/04, Cl2R1:44), (Cl2P7/04, Cl2R1:265), | |
| | PC (Cl2P7/04, Cl2R1:05), (Cl2P7/04, Cl2R1:01), (Cl2P7/04, Cl2R1:645) | |
| | CC A method of producing prenyl alcohol | |
| | PH Key Location/Qualifiers | |
| | FT CDS (1)..(1005). | |
| | | |
| FEATURES | Location/Qualifiers | |
| source | 1..1008 | |
| | /organism="Saccharomyces cerevisiae" | |
| | /mol_type="genomic DNA" | |

| | | | |
|----------------------------|-----|--|--|
| Db | | 961 TTACATGACGAATTGTTATATATATATATAGACCACTTATCCGAATTGTGA 1008 | |
| RESULT 5 | | | |
| BD170961 | | | |
| LOCUS | | 1008 bp DNA linear PAT 17-JAN-2003 | |
| DEFINITION | | Process for producing prenyl alcohol. | |
| ACCESSION | | BD170961 | |
| VERSION | | BD170961.1 GI:27876773 | |
| KEYWORDS | | WO 02053746-A/3. | |
| SOURCE | | Saccharomyces cerevisiae (baker's yeast) | |
| ORGANISM | | Saccharomyces cerevisiae | |
| REFERENCE | | Saccharomyces cerevisiae | |
| AUTHORS | | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | |
| TITLE | | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | |
| JOURNAL | | 1 (bases 1 to 1008) | |
| COMMENT | | Oto,C., Obata,S., Muramatsu,M., Nishi,K. and Totusuka,K. | |
| | | Process for producing prenyl alcohol | |
| | | Patent: WO 02053746-A 3 11-JUL-2002; | |
| | | TOYOTA MOTOR CORP,CHIKARA OTO,SHUSEI OBATA,MASAYOSHI MURAMATSU, | |
| | | KIYOHICO NISHI,KAZUHIKO TOTUSUKA | |
| | | OS Saccharomyces cerevisiae (yeast) | |
| | | PN WO 02053746-A/3 | |
| | | PD 11-JUL-2002 | |
| | | PF 28-DEC-2001 WO 2001JP011214 | |
| | | PR 28-DEC-2001 JP 00P 403067 | |
| | | PI CHIKARA OTO,SHUSEI OBATA,MASAYOSHI MURAMATSU, KIYOHICO NISHI, | |
| | | PI KAZUHIKO TOTUSUKA | |
| | | PC C12N15/52,C12P7/04,C12N1/19,C12N1/21 | |
| | | CC Process for producing prenyl alcohol | |
| | | FH Key Location/Qualifiers | |
| | | FT CDS (1)..(1005). | |
| FEATURES | | Location/Qualifiers | |
| source | | 1..1008 | |
| | | /organism="Saccharomyces cerevisiae" | |
| | | /mol_type="genomic DNA" | |
| | | /db_xref="taxon:4932" | |
| ORIGIN | | | |
| Query Match | | 64.2%; Score 1008; DB 6; Length 1008; | |
| Best Local Similarity | | 100.0%; Pred. No. 5.9e-195; | |
| Matches 1008; Conservative | | 0; Mismatches 0; Indels 0; Gaps 0; | |
| Qy | 301 | ATGAGGCCAAGATAGATGAGCTGATCAATAATGATCTCTTTGGTCCAGCCAAAATGAA 360 | |
| Db | 1 | ATGAGGCCAAGATAGATGAGCTGATCAATAATGATCTCTTTGGTCCAGCCAAAATGAA 60 | |
| Qy | 361 | AGCTTGATTTCAAAACCTTATATCATCTCTTTGAAACTGCGCAAGAACTTTAGACTA 420 | |
| Db | 61 | AGCTTGATTTCAAAACCTTATATCATCTCTTTGAAACTGCGCAAGAACTTTAGACTA 120 | |
| Qy | 421 | AAATTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGCTGGCCATGTT 480 | |
| Db | 121 | AAATTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGCTGGCCATGTT 180 | |
| Qy | 481 | TGCAAAATTTGATGAGCTCTTGCAATAATTCAGCCCTTTTAATTCGACGATATAGAGATAAT 540 | |
| Db | 181 | TGCAAAATTTGATGAGCTCTTGCAATAATTCAGCCCTTTTAATTCGACGATATAGAGATAAT 240 | |
| Qy | 541 | GCTCCCTTGAGAGGGGACAGCCACTTCTCAGTAAATCTTCGGTGTACCCCTCCACTATA 600 | |
| Db | 241 | GCTCCCTTGAGAGGGGACAGCCACTTCTCAGTAAATCTTCGGTGTACCCCTCCACTATA 300 | |
| Qy | 601 | AACACCCGCAATTAATATGATTTTCAGAGCCATGCACTTGTATCGAGCTTAACCAAAA 660 | |
| Db | 301 | AACACCCGCAATTAATATGATTTTCAGAGCCATGCACTTGTATCGAGCTTAACCAAAA 360 | |
| Qy | 661 | GAGCTTTGATCATTAATTTGATACGATTTTCAACGAAGAAATGATCAATCTACATAGG 720 | |
| Db | 361 | GAGCTTTGATCATTAATTTGATACGATTTTCAACGAAGAAATGATCAATCTACATAGG 420 | |
| Qy | 721 | GGACAAAGCTTTGGATATATATCTGGAGAGACTTTCTGCTGAAATCATCTACCTACAGAG 780 | |
| Db | 421 | GGACAAAGCTTTGGATATATATCTGGAGAGACTTTCTGCTGAAATCATCTACCTACAGAG 480 | |

```
QY 781 ATGTATTTGAATATGTTATGATTAATAAAGAGGGGCTTTTTCAGATTAACGTTGAGACTC 840
|
|
|
Db 481 ATGTATTTGAATATGTTATGATTAATAAAGAGGGGCTTTTTCAGATTAACGTTGAGACTC 540
|
|
|
QY 841 ATGGAAGCGCTGTCCTTCCTCCTCAGACCGGCAATCGTTGGTTCCTTTCATAAATCTT 900
|
|
|
Db 541 ATGGAAGCGCTGTCCTTCCTCCTCAGACCGGCAATCGTTGGTTCCTTTCATAAATCTT 600
|
|
|
QY 901 CTGGGTATTTATTCAGATTAGAGATGATTAATGAAATTCGAAATTCGAAATTCGAAATTC 960
|
|
|
Db 601 CTGGGTATTTATTCAGATTAGAGATGATTAATGAAATTCGAAATTCGAAATTCGAAATTC 660
|
|
|
QY 961 AGCGAAAAAGGCTTGTCTGAGGACATTAAGAGGGGAGGATTAATCTTTCCCATCGTCCAC 1020
|
|
|
Db 661 AGCGAAAAAGGCTTGTCTGAGGACATTAAGAGGGGAGGATTAATCTTTCCCATCGTCCAC 720
|
|
|
QY 1021 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACACAATGAAATTCGAAATTC 1080
|
|
|
Db 721 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACACAATGAAATTCGAAATTC 780
|
|
|
QY 1081 CTCCTGTTGAGGACAAAGTATAAAGATATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1140
|
|
|
Db 781 CTCCTGTTGAGGACAAAGTATAAAGTATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 840
|
|
|
QY 1141 GACACCAATTCATTTGGCTTACACCAAAATTTTAAATTAATTAATTAATTAATTAATTAAT 1200
|
|
|
Db 841 GACACCAATTCATTTGGCTTACACCAAAATTTTAAATTAATTAATTAATTAATTAATTAAT 900
|
|
|
QY 1201 AATGATTAAGAAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260
|
|
|
Db 901 AATGATTAAGAAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
|
|
|
QY 1261 TTACATGACGAAATTTGTTATATATATATATATATATATATATATATATATATATATATAT 1308
|
|
|
Db 961 TTACATGACGAAATTTGTTATATATATATATATATATATATATATATATATATATATATAT 1008

RESULT 6
LOCUS BD171079 1008 bp DNA linear PAT 17-JAN-2003
DEFINITION Process for producing prenyl alcohol.
ACCESSION BD171079
VERSION BD171079.1 GI:27876891
KEYWORDS WO 02053747-A/3.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 1008)
Oto,C. and Obata,S.
Process for producing prenyl alcohol
Patent: WO 02053747-A 3 11-JUL-2002;
TOYOFA MOTOR CORP,CHIKARA OTO,SHUSEI OBATA
OS Saccharomyces cerevisiae (yeast)
PN WO 02053747-A/3
PD 11-JUL-2002
PF 20-DEC-2001 WO 2001JP011215
PR 28-DEC-2000 JP 00P 401701.28-DEC-2000 JP 00P 403067 PR
18-SEP-2001 JP 01P 282978
PI CHIKARA OTO,SHUSEI OBATA
PC C12N15/52,C12P7/04,C12N1/19,C12N1/21
CC Process for producing prenyl alcohol
FH Key Location/Qualifiers
FT CDS (1)..(1005).
FEATURES
source
1..1008
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"

ORIGIN
Query Match 64.2%; Score 1008; DB 6; Length 1008;
```

```
Best Local Similarity 100.0%; Pred. No. 5.9e-195;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCTCTGTTTGGTCCAGCCAAATGAA 360
|
|
|
Db 1 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCTCTGTTTGGTCCAGCCAAATGAA 60
|
|
|
QY 361 AGCTTGATTTCAAAACCTTATATCAATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 420
|
|
|
Db 61 AGCTTGATTTCAAAACCTTATATCAATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 120
|
|
|
QY 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGGCCATAGTT 480
|
|
|
Db 121 AATTTAAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGGCCATAGTT 180
|
|
|
QY 481 TCGCAAAATTTGTGAGCTCTTTCGATTAATCCAGCCCTTTTAAATCGACGATATAGAGATTAAT 540
|
|
|
Db 181 TCGCAAAATTTGTGAGCTCTTTCGATTAATCCAGCCCTTTTAAATCGACGATATAGAGATTAAT 240
|
|
|
QY 541 GTCCTCTTGGAGAGGGGACAGACCACTTCTCACTTAATCTTTCGGTGTACCCCTCCACTATA 600
|
|
|
Db 241 GTCCTCTTGGAGAGGGGACAGACCACTTCTCACTTAATCTTTCGGTGTACCCCTCCACTATA 300
|
|
|
QY 601 AACACCGCAAAATTAATGCTATTTTCAGAGCCATGCAACTTGTATCGAGCTAACCAACAAA 660
|
|
|
Db 301 AACACCGCAAAATTAATGCTATTTTCAGAGCCATGCAACTTGTATCGAGCTAACCAACAAA 360
|
|
|
QY 661 GAGCCTTTGTATCATTAATTTGATTACGATTTTCAACGAGAAATTTGATCAATCTACATAGG 720
|
|
|
Db 361 GAGCCTTTGTATCATTAATTTGATTACGATTTTCAACGAGAAATTTGATCAATCTACATAGG 420
|
|
|
QY 721 GGACAAAGCCTTGGATATATATCTGGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAG 780
|
|
|
Db 421 GGACAAAGCCTTGGATATATATCTGGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAG 480
|
|
|
QY 781 ATGTATTTGAATATGTTATGATTAATAAAGAGGGGAGTATCTTTTCCCATCGTCCAC 840
|
|
|
Db 481 ATGTATTTGAATATGTTATGATTAATAAAGAGGGGAGTATCTTTTCCCATCGTCCAC 540
|
|
|
QY 841 ATGGAAGCGCTGTCCTTCCTCCTCAGACCGGCAATCGTTGGTTCCTTTCATAAATCTT 900
|
|
|
Db 541 ATGGAAGCGCTGTCCTTCCTCCTCAGACCGGCAATCGTTGGTTCCTTTCATAAATCTT 600
|
|
|
QY 901 CTGGGTATTTATTCAGATTAGAGATGATTAATGAAATTCGAAATTCGAAATTCGAAATTC 960
|
|
|
Db 601 CTGGGTATTTATTCAGATTAGAGATGATTAATGAAATTCGAAATTCGAAATTCGAAATTC 660
|
|
|
QY 961 AGCGAAAAAGGCTTGTCTGAGGACATTAAGAGGGGAGGATTAATCTTTCCCATCGTCCAC 1020
|
|
|
Db 661 AGCGAAAAAGGCTTGTCTGAGGACATTAAGAGGGGAGGATTAATCTTTCCCATCGTCCAC 720
|
|
|
QY 1021 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACACAATGAAATTCGAAATTC 1080
|
|
|
Db 721 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACACAATGAAATTCGAAATTC 780
|
|
|
QY 1081 CTCCTGTTGAGGACAAAGTATAAAGATATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1140
|
|
|
Db 781 CTCCTGTTGAGGACAAAGTATAAAGATATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 840
|
|
|
QY 1141 GACACCAATTCATTTGGCTTACACCAAAATTTTAAATTAATTAATTAATTAATTAATTAAT 1200
|
|
|
Db 841 GACACCAATTCATTTGGCTTACACCAAAATTTTAAATTAATTAATTAATTAATTAATTAAT 900
|
|
|
QY 1201 AATGATTAAGAAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260
|
|
|
Db 901 AATGATTAAGAAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960
|
|
|
QY 1261 TTACATGACGAAATTTGTTATATATATATATATATATATATATATATATATATATATATAT 1308
|
|
|
Db 961 TTACATGACGAAATTTGTTATATATATATATATATATATATATATATATATATATATATAT 1008
```

RESULT 7

| | | | | | |
|------------|---|---|-----|--------|-----------------|
| AY692852 | AY692852 | 1008 bp | DNA | linear | PLN 11-AUG-2004 |
| LOCUS | Saccharomyces cerevisiae clone FLH11088.01X YFL069C gene, complete cds. | | | | |
| DEFINITION | AY692852 | | | | |
| ACCESSION | AY692852.1 | GI:51013154 | | | |
| VERSION | Yeast ORF Project. | | | | |
| KEYWORDS | Saccharomyces cerevisiae (baker's yeast) | | | | |
| SOURCE | Saccharomyces cerevisiae | | | | |
| ORGANISM | Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | | | |
| REFERENCE | 1 (bases 1 to 1008) | | | | |
| AUTHORS | Marsischky, G., Rolfs, A., Richardson, A., Kane, M., Baqui, M., Taycher, E., Hu, Y., Vamberg, F., Weger, J., Kramer, J., Moreira, D., Kelley, F., Zuo, D., Raphael, J., Hogle, C., Jepson, D., Williamson, J., Camargo, A., Gonzaga, L., Vasconcelos, A.T., Simpson, A., Kolodner, R., Harlow, E. and LaBaer, J. | | | | |
| TITLE | Creation of the YFLEX clone resource: cloning of Saccharomyces cerevisiae ORFs in the Gateway recombinational cloning system | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 1008) | | | | |
| AUTHORS | Marsischky, G., Rolfs, A., Richardson, A., Kane, M., Baqui, M., Taycher, E., Hu, Y., Vamberg, F., Weger, J., Kramer, J., Moreira, D., Kelley, F., Zuo, D., Raphael, J., Hogle, C., Jepson, D., Williamson, J., Camargo, A., Gonzaga, L., Vasconcelos, A.T., Simpson, A., Kolodner, R., Harlow, E. and LaBaer, J. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (20-JUL-2004) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles st., Cambridge, MA 02141, USA | | | | |
| COMMENT | This clone is part of a collection of Saccharomyces cerevisiae full-length ORF clones generated by the Harvard Institute of Proteomics. Each CDS has been cloned with its native stop-codon. The CDS has been directionally cloned using the Gateway cloning system into the donor vectors pDONR 201 or pDONR 221. Additional sequences in the clone: 'TCAGCTGACAC' after the attL1 site and before the 'ATG' (from Research Genetics primers used to amplify the ORFs, including a Kozak consensus sequence); 'ATCCCGGAATGCCAG' after the stop codon and before the attL2 site (from the Research Genetics primers used to amplify the ORFs). | | | | |
| FEATURES | <p>Location/Qualifiers</p> <p>1..1008</p> <p>/organism="Saccharomyces cerevisiae"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:4932"</p> <p>/clone="FLH11088.01X"</p> <p>/lab_host="DH5alpha T1 resistant"</p> <p><1..>1008</p> <p>/product="YPL069C"</p> <p>1..1008</p> <p>/codon_start=1</p> <p>/product="YPL069C"</p> <p>/protein_id="AAT92871.1"</p> <p>/db_xref="GI:51013155"</p> <p>translation="MEAKIDELINNDPWSQNESLTFKYPNHLKPKGNRLNLIQINRVMLPKDQIAIVSQIVELLHNSLLDIEDNAPLRGOTTHLIFGVSTINTANYNYFRAMQLVSLQTTKEPYLNLIITIEEELNLRHGGLDIYWRDFPEIIPTEMYLNVMVNTKGGFRULRLRLMEALSPSHGSLVFPINLGLIYQIRDDYLNKDFQMSSEKGFADITTEKGLFFPIVHALNFTTKTGQEQHNEILRIILLRLTRSDKDKLKLQILFEDTNSLAYTKNFQNLVNMINKNDNENKYLPLDASHSDPATNLHDELLYIIDLHSE L"</p> | | | | |
| ORIGIN | <p>Query Match 64.2%; Score 1008; DB 8; Length 1008;</p> <p>Best Local Similarity 100.0%; Pred. No. 5.9e-195;</p> <p>Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p> | | | | |
| Qy | 301 | ATGAGGCCAGATAGATGAGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA | 360 | | |
| Db | 1 | ATGAGGCCAGATAGATGAGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA | 60 | | |
| Qy | 361 | AGCTTGATTTCAAAACCTTATAATCACATCTCTTTGAAACCTGGCAAGACTTTAGACTA | 420 | | |

| | | | |
|----|------|---|------|
| Db | 61 | AGCTTGATTTCAAAACCTTATAATCACATCTCTTTGAAACCTGGCAAGACTTTAGACTA | 120 |
| Qy | 421 | AATTTAAATAGTTCAAAATTAACAGAGTTTATGAATTTGGCCCAAGACCAGCTGGCCATAGTT | 480 |
| Db | 121 | AATTTAAATAGTTCAAAATTAACAGAGTTTATGAATTTGGCCCAAGACCAGCTGGCCATAGTT | 180 |
| Qy | 481 | TCGCAAAATTTGTTGAGCTCTTGATATAATTCACGCCCTTTTAATTCAGATATAGAAGTAAT | 540 |
| Db | 181 | TCGCAAAATTTGTTGAGCTCTTGATATAATTCACGCCCTTTTAATTCAGATATAGAAGTAAT | 240 |
| Qy | 541 | GCTCCCTTGAGAGAGGGGACAGACCTTCTCACTTAATCTTCGCTGCTGCTGCTCCACTATA | 600 |
| Db | 241 | GCTCCCTTGAGAGAGGGGACAGACCTTCTCACTTAATCTTCGCTGCTGCTGCTCCACTATA | 300 |
| Qy | 601 | AACACGGCAAAATTAATATGATTTTCAGAGCCATGCAACTTTGTATCGCAGCTAACCAAAA | 660 |
| Db | 301 | AACACGGCAAAATTAATATGATTTTCAGAGCCATGCAACTTTGTATCGCAGCTAACCAAAA | 360 |
| Qy | 661 | GAGCCTTTGATCATTAATTTGATTTTACGATTTTCAACGAAAGTAATTCATCTACATAGG | 720 |
| Db | 361 | GAGCCTTTGATCATTAATTTGATTTTACGATTTTCAACGAAAGTAATTCATCTACATAGG | 420 |
| Qy | 721 | GGACAGGCTTGGATATATCTGAGAGACTTTCTGCTGAAATCATCTACTCTACAGGAG | 780 |
| Db | 421 | GGACAGGCTTGGATATATCTGAGAGACTTTCTGCTGAAATCATCTACTCTACAGGAG | 480 |
| Qy | 781 | ATGTATTTCATATATGTTGATTAATAAACAAGCGGCTTTTTCAGATTAAAGTTGAGACTC | 840 |
| Db | 481 | ATGTATTTCATATATGTTGATTAATAAACAAGCGGCTTTTTCAGATTAAAGTTGAGACTC | 540 |
| Qy | 841 | ATGGAAGCGCTGTCCTTCTCCACACGAGGCAATCGTTGGTTCCTTTTCATAAATCTT | 900 |
| Db | 541 | ATGGAAGCGCTGTCCTTCTCCACACGAGGCAATCGTTGGTTCCTTTTCATAAATCTT | 600 |
| Qy | 901 | CTGGGTATTTATTTATCAGATTAGAGATGATTAATTCGAATTTTGAAGATTTCCAAATGTC | 960 |
| Db | 601 | CTGGGTATTTATTTATCAGATTAGAGATGATTAATTCGAATTTTGAAGATTTCCAAATGTC | 660 |
| Qy | 961 | AGCGAAAAGGCTTTGCTCAGAGCATTACAGAGGGGAAAGTTATCTTTTCCATCGTCCAC | 1020 |
| Db | 661 | AGCGAAAAGGCTTTGCTCAGAGCATTACAGAGGGGAAAGTTATCTTTTCCATCGTCCAC | 720 |
| Qy | 1021 | GCCTTTAACTTCTCTAAAACGAAAGGTCAAACTGAGCAACAATGAAATTTCTAAGAATT | 1080 |
| Db | 721 | GCCTTTAACTTCTCTAAAACGAAAGGTCAAACTGAGCAACAATGAAATTTCTAAGAATT | 780 |
| Qy | 1081 | CTCCTGTTGAGGACAAAGTGATAAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC | 1140 |
| Db | 781 | CTCCTGTTGAGGACAAAGTGATAAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC | 840 |
| Qy | 1141 | GACACCAATTCATTGGCCTACACCAAAATTTTATTAAATCAATAGTGAATATGATAAAA | 1200 |
| Db | 841 | GACACCAATTCATTGGCCTACACCAAAATTTTATTAAATCAATAGTGAATATGATAAAA | 900 |
| Qy | 1201 | AATGATTAATGAAAATAGTATTTACCTGATTTGGCTTCGCATTCGACACCGCCACCAAT | 1260 |
| Db | 901 | AATGATTAATGAAAATAGTATTTACCTGATTTGGCTTCGCATTCGACACCGCCACCAAT | 960 |
| Qy | 1261 | TTATCATGACCAATTTGTTTATATATAATAGACCACTTATCCGAATTTGCA 1308 | |
| Db | 961 | TTATCATGACCAATTTGTTTATATATAATAGACCACTTATCCGAATTTGCA 1008 | |

| | | | | | | |
|------------|---------------|------------------------------------|---------|-----|--------|-----------------|
| RESULT 8 | AR071896 | Sequence 1 from patent US 5912154. | 1005 bp | DNA | linear | PAT 18-FEB-2000 |
| LOCUS | AR071896 | Sequence 1 from patent US 5912154. | | | | |
| DEFINITION | AR071896 | | | | | |
| ACCESSION | AR071896 | | | | | |
| VERSION | AR071896.1 | GI:7222784 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | Unknown. | | | | | |
| ORGANISM | Unclassified. | | | | | |

REFERENCE 1 (bases 1 to 1005)
AUTHORS Ferro-Novick, S. and Jiang, Y.
TITLE Geranylgeranyl diphosphate synthase proteins, nucleic acid molecules and uses thereof
JOURNAL Patent: US 5912154-A 1 15-JUN-1999;
FEATURES location/Qualifiers
source I. .1005
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 64.1%; Score 1005; DB 6; Length 1005;
Best Local Similarity 100.0%; Pred. No. 2.4e-194;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCCTGTTGGTCCAGCCCAAAATGAA 360
DB 1 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCCTGTTGGTCCAGCCCAAAATGAA 60
QY 361 AGCTTGATTTCAAAACCTTTATATACATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 420
DB 61 AGCTTGATTTCAAAACCTTTATATACATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 120
QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACAGCTGCCCATAGTT 480
DB 121 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACAGCTGCCCATAGTT 180
QY 481 TCGCAAAATGTTGAGCTCTTGCAATATCCAGCCCTTTTAAATCGAGATATAGAAATAT 540
DB 181 TCGCAAAATGTTGAGCTCTTGCAATATCCAGCCCTTTTAAATCGAGATATAGAAATAT 240
QY 541 GTCCTCTTGAGAAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACTCCCTCCACTATA 600
DB 241 GTCCTCTTGAGAAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACTCCCTCCACTATA 300
QY 601 AACACCGCAAAATATATATGTTTACAGAGCATGCACTTGTATCGCAGTAAACCAAAAA 660
DB 301 AACACCGCAAAATATATGTTTACAGAGCATGCACTTGTATCGCAGTAAACCAAAAA 360
QY 661 GAGCCTTTGATATCAATTTGATACGATTTTCAACGAGAAATTTGATCGCAGTAAACCAAAAA 720
DB 361 GAGCCTTTGATATCAATTTGATACGATTTTCAACGAGAAATTTGATCGCAGTAAACCAAAAA 420
QY 721 GGA CAAGCCTTGATATATATGATGAGAGACTTCTGCTGAAATCATACCTACTCAGGAG 780
DB 421 GGA CAAGCCTTGATATATATGATGAGAGACTTCTGCTGAAATCATACCTACTCAGGAG 480
QY 781 ATGTATTTGAATATGTTTATCAATAAACAAGCGGCTTTTTCAGATTAACGTTGAGATC 840
DB 481 ATGTATTTGAATATGTTTATCAATAAACAAGCGGCTTTTTCAGATTAACGTTGAGATC 540
QY 841 ATGAAGCGCTGTCTCTCTCTCAACACGCGCATTCGTTGGTTCTCTTTCATAAATCTT 900
DB 541 ATGAAGCGCTGTCTCTCTCTCAACACGCGCATTCGTTGGTTCTCTTTCATAAATCTT 600
QY 901 CTGGGTATTTATTCAGATTTAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTC 960
DB 601 CTGGGTATTTATTCAGATTTAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTC 660
QY 961 AGCCAAAAGGCTTTGCTGAGGACATTTACAGAGGGAGTTATCTTTTCCCATCGTCCAC 1020
DB 661 AGCCAAAAGGCTTTGCTGAGGACATTTACAGAGGGAGTTATCTTTTCCCATCGTCCAC 720
QY 1021 GCCCTTAATCTTCACTAAACCAAGGTTCAAACTGAGCAACAATGAAATTTCTAAGATTT 1080
DB 721 GCCCTTAATCTTCACTAAACCAAGGTTCAAACTGAGCAACAATGAAATTTCTAAGATTT 780
QY 1081 CTCCTGTGAGGACAAAGTATATAAATACTAAGCTGATTTCAAAATCTGAAATTC 1140
DB 781 CTCCTGTGAGGACAAAGTATATAAATACTAAGCTGATTTCAAAATCTGAAATTC 840
QY 1141 GACACCAATTCATTGGCTCACCAAAAATTTTAAATCAATTAAGTGAATATGATATAA 1200

DB 841 GACACCAATTCATTGCGCTACACCAAAATTTTATTAATCAATTAAGTGAATGATAAA 900
QY 1201 AATGATATGAAAAATAAGTATTACTGATTTGGCTTGCATTTCCGACACCGCCACCAAT 1260
DB 901 AATGATATGAAAAATAAGTATTACTGATTTGGCTTGCATTTCCGACACCGCCACCAAT 960
QY 1261 TTACATGAGGAATTTGTTATATATATAGACCACTTATCCGAATTG 1305
DB 961 TTACATGAGGAATTTGTTATATATATAGACCACTTATCCGAATTG 1005
RESULT 9
CR380954.04
WPCOMMENT
Sequence split into 11 fragments LOCUS CR380954 Accession CR380954
Fragment Name Begin End
CR380954_00 1 110000
CR380954_01 100001 210000
CR380954_02 200001 310000
CR380954_03 300001 410000
CR380954_04 400001 510000
CR380954_05 500001 610000
CR380954_06 600001 710000
CR380954_07 700001 810000
CR380954_08 800001 910000
CR380954_09 900001 1010000
CR380954_10 1000001 1050361
Continuation (S of 11) of CR380954 from base 400001 (CR380954 Candida glabrata strain CBK
Query Match 17.7%; Score 277.2; DB 8; Length 110000;
Best Local Similarity 53.8%; Pred. No. 1.6e-46;
Matches 623; Conservative 0; Mismatches 523; Indels 12; Gaps 2;
QY 337 CTTGTTTGGTCCAGCCAAATGAAAGCTTGATTTCAAACCTTATAATACATCATTCGCTTTG 396
DB 108599 CCAGCATGTCACAAATCAGATGAAAAATTCATATAATCAACCATACATACATTCGCA 108658
QY 397 AAACCTGCGAAGAACTTTAGACTAAATTTAATAGTTCAAATTAACAGAGTTATGAATTTG 456
DB 108659 AACCGGGAAGAACTTCAGAGCCAAATTAATACATCTTTTAAACCACTTCTATTGCTA 108718
QY 457 CCCAAAGACCACTGGCCATAGTTTGCATAATGTTGAGCTCTTGCATAATTTCCAGCCTT 516
DB 108719 GATGAGATATTTATACAAACACTCACTCAATTTGAGAGATCTCCCAATTTCAAGTTG 108778
QY 517 TTAATCGAGATATAGAGATAATGCTCCCTTGAGAGGGGACAGACCACTTCTCAGCTTA 576
DB 108779 ATTATCGATGACATTTGAAGATACTCACAGTTGAGGGGTGAGTAGTCGCAATCATATG 108838
QY 577 ATCTTCGGTGTACCCTCCACTATATAACACCGCAATTTATATGTTTTCAGAGCATGCAA 636
DB 108839 CTCATGGTGTACCCTATGTCATTAATTAATCTGCAAACTACATGTTATTTGTAGCAATG 108898
QY 637 CTTCATCGCAGCTTAACACCAAAAGAGCCTTTGTATCATATAATTTGATTACGATTTTCAAC 696
DB 108899 TTACTTAGAACCTGTCTCATGATGAGTAACTTCAATGATCTAATGAAATATTTAAC 108958
QY 697 GAAGAAATGATCAATCTTACATAGGGGCAAGGCTTGGATATATATCTGGAGAGACTTTCTG 756
DB 108959 GAGGAACCTTCTATATCTTTCAGAGGCCAAGGATTTGATATATATCTGGCGTACTCTTA 109018
QY 757 CCTCAATCATACTACTCTAGGAGATGTTTGAATATGTTTGAATATAAACAAGCGGCG 816
DB 109019 CCTAAGGTTGTACCAACAGAGGAATGTTATTTCAACATGTTTATGAATAAACTGGTGA 109078
QY 817 CTTTTCAGATTAAGTTGAGACTCATCGAAGCGGTCTCTCTTCTCCACACCGGCGCAT 876
DB 109079 CTCTTCAGATTAAGCTTAAATTCATGGAACGCTTAAGCGAATCTGGCAGGCTCAGAAA 109138
QY 877 TCGTTGGTTCCTTTCAATAATCTTCTGGGTATTTATTTATCAGATTTAGATGATTTACTTGT 936
DB 109139 TCATTAGTTCCATTGGGAATTTTACTAGGAATTTATATATCAAGTTAGAGACGATTTACTTGT 109198

| | | | |
|----|--------|----------------|--------|
| QY | 937 | AATTGAAAGATTTC | 996 |
| DB | 109199 | AATCTTACTGACTC | 109258 |
| QY | 997 | AAGTTTATCTTTT | 1056 |
| DB | 109259 | AAATTAATCTTTT | 109312 |
| QY | 1057 | CAACCAATGAAT | 1116 |
| DB | 109313 | AACAACACATG | 109372 |
| QY | 1117 | AGCTGATTC | 1176 |
| DB | 109373 | AATGCCATACG | 109432 |
| QY | 1177 | AATCAATAGT | 1236 |
| DB | 109433 | AAAAAATGAC | 109492 |
| QY | 1237 | TGCAATTCG | 1296 |
| DB | 109493 | ACCAATAGC | 109546 |
| QY | 1297 | TCCGAATTC | 1356 |
| DB | 109547 | GCTAATG | 109606 |
| QY | 1357 | CCTTCTCTC | 1416 |
| DB | 109607 | CCTGACCTT | 109666 |
| QY | 1417 | ATCTTTTG | 1476 |
| DB | 109667 | TTGATGAT | 109726 |
| QY | 1477 | TAGCCCTT | 1494 |
| DB | 109727 | CATAAATA | 109744 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|----|
| RESULT 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | </ |
|-----------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|----|

| | | | |
|----|------|------------------|------|
| DB | 8719 | GATGAGAAATATTTAT | 8778 |
| QY | 517 | TTAATCGACGATAT | 576 |
| DB | 8779 | ATTATCGATGACAT | 8838 |
| QY | 577 | ATCTTCGGTGTAC | 636 |
| DB | 8839 | CTCTATGGTGTAC | 8898 |
| QY | 637 | CTTGATTCGACG | 696 |
| DB | 8899 | TTACTTAGAAAC | 8958 |
| QY | 697 | GAAGAAATGATCA | 756 |
| DB | 8959 | GAGGAACCTTCT | 9018 |
| QY | 757 | CCTGAAATCAT | 816 |
| DB | 9019 | CCTAAGGTTGT | 9078 |
| QY | 817 | CTTTTCAGATTA | 876 |
| DB | 9079 | CTCTTCAGATTA | 9138 |
| QY | 877 | TCGTTGGTTCC | 936 |
| DB | 9139 | TCATTAGTTCC | 9198 |
| QY | 937 | AATTTGAAAG | 996 |
| DB | 9199 | AATCTTACTG | 9258 |
| QY | 997 | AGTTTATCTTT | 1056 |
| DB | 9259 | AAATTTATCT | 9312 |
| QY | 1057 | CAACACAATGA | 1116 |
| DB | 9313 | AACAACACAT | 9372 |
| QY | 1117 | AAGCTGATTA | 1176 |
| DB | 9373 | AATGCCATAG | 9432 |
| QY | 1177 | AATCAATTAG | 1236 |
| DB | 9433 | AAAAAATTG | 9492 |
| QY | 1237 | TCCGATTTCC | 1296 |
| DB | 9493 | ACCAATAGC | 9546 |
| QY | 1297 | TCCGAATG | 1356 |
| DB | 9547 | GCTAATG | 9606 |
| QY | 1357 | CTTCTCTCT | 1416 |
| DB | 9607 | CTTGACCTT | 9666 |
| QY | 1417 | ATCTTTTG | 1476 |
| DB | 9667 | TTGATGAT | 9726 |
| QY | 1477 | TAGCCCTT | 1494 |
| DB | 9727 | CATAAATA | 9744 |

SCU41849 35868 bp DNA linear PLN 02-SEP-1997
Saccharomyces cerevisiae chromosome XVI, left arm, cosmid 8204.
ACCESSION U41849 U00094
VERSION U41849.1 GI:1147608
KEYWORDS SEC16; MOT1; ATP4; S.cerevisiae 60S ribosomal protein L21B homolog; S.cerevisiae Ybr197p homolog; GPI2; GCR1; YTA6.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 35868)
AUTHORS Baker, H.V.
TITLE Glycolytic gene expression in Saccharomyces cerevisiae: nucleotide sequence of GCR1, null mutants, and evidence for expression
JOURNAL Mol. Cell. Biol. 6 (11), 3774-3784 (1986)
MEDLINE 87089718
PUBMED 3025612
REFERENCE 2 (bases 1 to 35868)
AUTHORS Velours, J., Durrans, P., Aigle, M. and Guerin, B.
TITLE ATP4, the structural gene for Yeast FOF1 ATPase subunit 4
JOURNAL Eur. J. Biochem. 170 (3), 637-642 (1988)
MEDLINE 88111651
PUBMED 2892678
REFERENCE 3 (bases 1 to 35868)
AUTHORS Davis, J.L., Kunisawa, R. and Thorner, J.
TITLE A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast Saccharomyces cerevisiae
JOURNAL Mol. Cell. Biol. 12 (4), 1879-1892 (1992)
MEDLINE 92195335
PUBMED 1312673
REFERENCE 4 (bases 1 to 35868)
AUTHORS Schnall, R., Mannhaupt, G., Stucka, R., Tauer, R., Ehmlé, S., Schwarze, C., Vetter, I. and Feldmann, H.
TITLE Identification of a set of yeast genes coding for a novel family of putative ATPases with high similarity to constituents of the 26S protease complex
JOURNAL Yeast 10 (9), 1141-1155 (1994)
MEDLINE 95274317
PUBMED 7754704
REFERENCE 5 (bases 1 to 35868)
AUTHORS Leidich, S.D., Kostova, Z., Latek, R.R., Costello, L.C., Drapp, D.A., Gray, W., Fassler, J.S. and Orlean, P.
TITLE Temperature-sensitive yeast GPI anchoring mutants gpi2 and gpi3 are defective in the synthesis of N-acetylglucosaminyl phosphatidylinositol. Cloning of the GPI2 gene
JOURNAL J. Biol. Chem. 270 (22), 13029-13035 (1995)
MEDLINE 95286583
PUBMED 7768836
REFERENCE 6 (bases 1 to 35868)
AUTHORS Espenshade, P., Gimeno, R.E., Holzmacher, E., Teung, P. and Kaiser, C.A.
TITLE Yeast SEC16 gene encodes a multidomain vesicle coat protein that interacts with Sec23p
JOURNAL J. Cell Biol. 131 (2), 311-324 (1995)
MEDLINE 96017704
PUBMED 7593161
REFERENCE 7 (bases 1 to 35868)
AUTHORS Bussey, H., Storms, R.K., Ahmed, A., Albertmann, K., Allen, E., Ansoorge, W., Araujo, R., Aparicio, A., Barrell, B., Badcock, K., Benes, V., Botstein, D., Bowman, S., Bruckner, M., Carpenter, J., Cherry, J.M., Chung, F., Churcher, C., Coster, F., Davis, K., Davis, R.W., Dietrich, F.S., Delius, H., Dipalo, T., Dubois, E., Dusterhoft, A., Duncan, M., Floeth, H., Fortin, N., Friesen, J.D., Fritz, C., Goffeau, A., Hall, J., Hebling, U., Humann, K., Hilbert, H., Hillier, L., Hunkle-Smith, S., Hyman, R., Johnston, M., Kalman, S., Kleene, K., Komp, C., Kurdi, O., Lashkari, D., Lew, H., Lin, A., Lin, D., Louis, E.J., Marathe, R., Messing, F., Mewes, H.W., Mirti, P., Moestl, D., Muller-Auer, S., Namath, A., Nentwich, U., Oefner, P., Pearson, D., Petel, F.X., Pohl, T.M., Purnelle, D., Schafer, M., Scharfe, M., Scherens, B., Schramm, S., Schroeder, M., Sedgwick, A.M., Tettelin, H., Urrestarazu, L.A., Ushinsky, S., Vierendeels, F., Vissers, S., Voss, H., Walsh, S.V., Wambutt, R., Wang, Y., Wedler, E., Wedler, H., Winnett, S., Zhong, W.W., Zollner, A., Vo, D.H. and Hani, J.
TITLE The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI

Nature 387 (6632 Suppl.), 103-105 (1997)
97313271
9169875
8 (bases 1 to 35868)
Hall, J., Ahmed, A., Bussey, H., Fortin, N., Friesen, J.D., Storms, R.K., Vo, D.H., Wang, Y. and Winnett, E.
The sequence of Saccharomyces cerevisiae chromosome XVI left arm
Unpublished
9 (bases 1 to 35868)
Bussey, H.
Direct Submission
Submitted (08-AUG-1995) Howard Bussey, McGill University, Biology, 1205 Dr. Penfield Ave., Montreal, Quebec H3A 1B1, Canada
10 (bases 1 to 35868)
Jia, Y. and Cherry, J.M.
Direct Submission
Submitted (02-SEP-1997) Department of Genetics, Stanford University, Saccharomyces Genome Database, Stanford, CA 94305-5120, USA
Curated by:
Saccharomyces Genome Database
URL: <http://genome-www.stanford.edu/>
e-mail: yeast-curator@genome.stanford.edu
All CDS of 100 or more codons have been analysed; CDS that are overlapped by larger CDS are not included; this cosmid is overlapped on the right with cosmid 8460, and on the left by 8047.
Location/Qualifiers
1..35868
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="XVI"
/map="XVI; left arm"
/note="cosmid 8204"
1..50
/note="overlaps with cosmid 8047."
51..6638
/gene="SEC16"
51..6638
/gene="SEC16"
/note="YPI085W; similar to Sec16p, a multidomain vesicle coat protein encoded by GenBank Accession number U23819."
codon_start=1
/product="Sec16p"
/protein_id="AAB68254.1"
/db_xref="GI:1147609"
/translation="MTPEAKRRKQKYLKQKAAEKAAASHSEPELEPESTINSS
FNDSVNRSTSDASKSDVPFVSSTNISPANETQLFIPDTQLHLHLNDSQDHDIT
ADNDLPDSIVHSDSVITQTPAMSOEYBETAHLSSRNFLSDVAVAGELHNNHTQ
KIAVSAVEDSFNEEGENHDSIISSINDATPSQYHNFLPSDGNLISPLISSGDTPT
HNVLPTKNEINDDEYCNDEKISLANNVLPDELSKEDEBRLEKLETHVSTEKKQDI
ADQETAEFLNSTSEPSKIRNSGDDTSMFLFODDESOKVPEEDVKQPHNENTNN
TGKNIENSOKLMGEGNHKPLSAEADIIEFGKDIQDAEDLFTQSSGDLGELVPWS
TDKNADVTSQKQEDLFAAGNDEKLPWESDGEVSSGKTSNMTSTKIAEQF
SFLEDDLLDDDDTSSFLASSEEDTVENTDNTNLTSPVEEKKASRYKPIIEEAGF
RQGVHFTNTGTVTPOQFGLTKGTQNOQVSPNIVSPKPVVKDRNSFKINE
EKKSDAYDFLEIISSEKKGAKPVAPTORGSGNSPSSLDKPIQSRKGSNNEN
RPVPLGTQEPSSRTNSAISQSPVNAFPNPKYIQQLOQAPQSGMPLPNTNIPP
ALKVETTVSAPPISRRGVNSVSSAGARHATQYGLANGVPPVPGYQOATNIPPT
ANKIAPVSTVQKQYPSVVQNLGASAVNTPNFKTHRGHTSSISSTPNQNEHASRY
APNYQSQYPTTSQPVGVAGNSYSQSTRSSVAVPMFQAOTASIQPHNATSVTA
GILPLAPRLDPLQALNLPASNTIANSIPLANPLAENTLPIIITHRATSVTA
PPROENKPIKDNEALRRQFPIFHSAAKVYAVPPIQSQYMISSISSIVQIEKVT
PIDQILKNDMLKSPFGPLGSAKLLKDKLTKWMTETIKSINSESSDMDTIWOLEMK
LNDKWNWKNISKLLNSDELLMLYSOPFPNGDMI PNAYRLDINQKRVLAFLQENHD
EALRLSKEDYAIALLVGLSMGKDRWSEVIQKLYEGFTAGPNDQKELAHFLILFIQ
VFGNSKMAIKTAAASTFLIIGNVPLSNPWNADSVIFESIGNMNTFELWIEIY
FGIPLTKKGLTAAASTFLIIGNVPLSNPWNADSVIFESIGNMNTFELWIEIY
IFSYDPKFKGFSILQKLYHALLQKGLNSLGTXYTLSSVRKLPKQILITNL
TRLSSEVSRLESNTGNLAKPLSSVWGLDKSNFKYIGGDDIDALNKKNDKKCVFD

GETPGSSANSSTVDLTOTFTPQAOVTSQSVYDVTALLHNAHNPVSHSVLHSPKSNVS
KGLVBNANPYTHRIGDLSGSSFORHNTQPAABEQMASLRVTRDQHTNEKALSKSQ
ILEKSTAYTPQFGQNHVMEKSNVPSLFPADFPAPPLGTGTVPNSVSSPDLVRR
SIITSGSEFLPPKIGVPTKANSQSLMSPSVEALPDPVQVQVHETGYNDGNGKH
SOKSMPEDESHTSNDNADQNTLKDADVTDTMDIEGPFNDKNLLPMEPNQPT
SVNFIQIOTSDIOPTLOTNVVRGTDAKSMENSLPSIENERSSEEQPENISKASSA
VLPTGGIISLENERPLTODENSISETVQSTYLPAGSTISMEAKPIISOVDVYRNVNKA
KLVQHMAPPKKSTADATKMNYSPPVQSTAAASADGDESTILKTSFPAIYRNVQHAAS
NPSQFPIVQNETASFELSSQSNVASENRSFPIKKAIEVVEKDTFQPTIR
KASTNQKFLPESDADQNDVIEDESDDDNMSTDEAKRKEKNVNMKEKPSN
KOIDDKSGFWGLKDTGDKVYAKLGHQNTLYDEKLEKRWKNDKPEBQKXIE
SSAPPPPTTKKGGQPKTKPSGPINNLSLPVPHATSVIPNNPIICEPIPIKTSPT
GNBNNSPSPSPISRIKSVNLTSKANGLLDLSLAGGPKPKPASTERKKKTARGVNV
MNIQ"
/gene .9556
/gene="BRO1"
/gene="BRO1"
/gene="BRO1"
/note="YPL084W; BCK1-like resistance to osmotic shock"
/codon_start=1
/evidence=experimental
/product="Bro1p"
/protein_id="AAB68255.1"
/db_xref="GI:1147610"
/translation="MKPFLFDLKLKDTKELDKWKKGLSSYLKKSYSQSWRTFFYDEKAT
SELHLRNANGELAPSSLSSEONLKYVSFLEHLYPRLGSKGSLKMDFTWDAEYSSA
QKGLKTQHLAFKSCCTLFNTAVITQIARENINEDYKNSIANLTAKASCEYLSN
FLNPSVDLSQNETFLANICHAEOLEFLKILNDQIISQKTLISKLRSATCNLFQ
KCHDMKETDDVAIVGEFTKWTFTVKLHFYKSLGAYYHGLEENVRGAEIAPLD
FSMQQLISLPLTWIYVEFDKFKETLEKQKELIKONDFIYHESVAVVQVDSIK
ALDAIKSPTEKILRPMODVANKCDLSYRGIIPLDVYKESYSEBEKATLARKQVEE
TETANLESFTEFTNRLLSLEKQFSDGNLFSNTDQGLMRDQIOTWCKFIOTN
EFDIDQNKVILFKKQILSILSALPNQKENVTKLKSLVAASNDEKLKACVPH
IYVINDLNGKIKWDFBNRTPPQSLDDIDDTKNDKILKLLQKHGAHDLRLT
KERSNLSELREINDDITKLLINKGSDVELKDLFEVELEKEPESLSTRATYI
KQSMDDDKAKLDEIFHLNFKDKSGBEKLEDRKFNFDKLEAVKSFIPASDLP
KGTEFYDLSFNMRDLAERVRVAKOTEDSTANSAPPPLPLDLSKSVGSGPPLPKQS
AAQSLRQGLNLCQPNLKIISAGSDLPQGGIIPRTYEASPYATPTMAAPVPPPK
QSQDMYDLRRKAVENRELOENFTSYFNRSVFDENYSKYSS"
/gene
/gene="SEN54"
/gene="SEN54"
/gene="SEN54"
/note="YPL083C"
/codon_start=1
/product="Sen54p: tRNA splicing endonuclease 54kD subunit"
/protein_id="AAB68256.1"
/db_xref="GI:1147611"
/translations="MOPAGKKTQVTTNSNPFEEEEEEELQDWSQSLASLVSKNAA
LSLPKRGKDYEPDNTLOLLIYNASKAMFDITSDISIRGTTVKSEVRGYVVPKHQA
VLLKPGSGFMQTRADSTGELMLDPEFVYLAERGTLIPYVLEAGSNKSKHETEI
LUSMEDILSFSQQEMDQIFVFAHLKRLGFIKPSNQEAARKTSFPPLKQKRSNLQA
ITWRLSLFKIQLSIFSPGFKWNPFFKRYTSPQYQLNRLVRSVAVPKNKKEL
LDAQSDREFQKVDIPLTFKVKWPKPHSKKRDPLDPQVVFVYKNDLQHFPTYPEL

Query Match 16.6% Score 261; DB 8; Length 35868;
Best Local Similarity 100.0%; Pred. No. 3.7e-43;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1309 AATAAATTGATCAATCAAAATAGTGGAGGAGATAGTCAGAAATAAAGCCTTCTCTCCTC 1368
Db 35868 AATAAATTGATCAATCAAAATAGTGGAGGAGATAGTCAGAAATAAAGCCTTCTCTCCTC 35809
QY 1369 CTCTTTGCGATCTATACATGATTTTATATATAGCTTTCATTCGATCATCTTTTGATAT 1428
Db 35808 CTCTTTGCGATCTATACATGATTTTATATATAGCTTTCATTCGATCATCTTTTGATAT 35749
QY 1429 ATCTCAAAAAGATCTCTAGTTTCGAAATAGTCAAAATCTTCAAAATTTATAGCCCTTTATAT 1488
Db 35748 ATCTCAAAAAGATCTCTAGTTTCGAAATAGTCAAAATCTTCAAAATTTATAGCCCTTTATAT 35689
QY 1489 TTTTTCACGATTTCTGAAATCTCTTTTATATCAGCACCGTTAAATGCTAGCGGTTACTGTC 1548

Db 35688 TTTTTCACGATTTCTGAAATCTCTTTTATCAGCACCGTTAAATGCTAGCGGTTACTGTC 35629
QY 1549 AATATCGCCGGTAAATTCGCGA 1569
Db 35628 AATATCGCCGGTAAATTCGCGA 35608
RESULT 12
AE016894/c 300316 bp DNA linear PLN 14-APR-2004
LOCUS Erethothecium gossypii chromosome V, section 1 of 5 of the complete
DEFINITION sequence.
ACCESSION AE016894 AE016818
VERSION AE016894.1 GI:44983156
KEYWORDS Erethothecium gossypii (Ashbya gossypii)
SOURCE Erethothecium gossypii
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Erethothecium.
REFERENCE 1 (bases 1 to 300316)
AUTHORS Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K.,
Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R.A.,
Flavier, A., Gaffney, T.D. and Philippsen, P.
TITLE The Ashbya gossypii genome as a tool for mapping the ancient
Saccharomyces cerevisiae genome
JOURNAL Science 304 (5668), 304-307 (2004)
PUBMED 15001715
REFERENCE 2 (bases 1 to 300316)
AUTHORS Gates, K., Dietrich, F.S., Brachat, S., Voegeli, S.E., Lerch, A.,
Philippsen, P. and Gaffney, T.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2002) Department of Ashbya Genomics, Syngenta,
Research Triangle Park, NC 27709, USA
COMMENT This is low coverage sequence generated to identify the complete
set of genes and the gene order on this chromosome. Regions of low
quality are identified. Before doing extensive work on any gene
identified here the sequence should be confirmed.
FEATURES
1. 300316 Location/Qualifiers
/organism="Erethothecium gossypii"
/mol_type="genomic DNA"
/strain="ATCC 10895"
/db_xref="ATCC:10895"
/db_xref="taxon:33169"
/chromosome="V"
/note="synonym: Ashbya gossypii"
4858. .32110
/organism="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone BAG1539"
62601. .127208
/organism="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone BAG1581"
103331. .181613
/organism="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone BAG1047"
169336. .224489
/organism="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone BAG1442"
210657. .270834
/organism="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone BAG1767"
233952. .295486
/organism="Erethothecium gossypii"

| | | | | | |
|-----------------------|---|--|--------|--------|-----------------|
| Db | 189975 | TGTTCAACGGGTTCTATGATGCTTTTCGGAGCGCAGGTCGCGCGGTACACAGATTGGTGG | 189916 | | |
| QY | 494 | AGCTCTTGATTAATCCAGCCCTTTTAATCGACGATATAGAGATAATGCTCCCTTGAGAA | 553 | | |
| Db | 189915 | AGATGTTGATGCGGAGCCTTGTGATCGACGATATCGACAACTCTGCATGAGAGAC | 189856 | | |
| QY | 554 | GGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATAAACACCGGAATT | 613 | | |
| Db | 189855 | GTGGCGTCCCTGCTGCCACGTTGATATGGCTCGCCCATGACGATCAACACTCGGAATT | 189796 | | |
| QY | 614 | ATATGATTTTCAGAGCCATGCACTTGTATCGCAGCTAACCAACAAA-----AGAGCCTT | 667 | | |
| Db | 189795 | ACATGATTTTGTGTCACATGTCGCTCGCGCAGCTGGCGCTCAGCGTCCGGCGGCC | 189736 | | |
| QY | 668 | TGTATCATATTTGATTAACGATTTTCAAGAAATGATCAATCTACATAGGGGACAG | 727 | | |
| Db | 189735 | CGCTGAGGATCTGCTGAAGTGTTCACAGGAGATGATGATCTGCTCGGGGACAG | 189676 | | |
| QY | 728 | GCTTGGATATATCTGGAGAGACTTTCCTGCTGAATCATACTACTCAGGAGATGATT | 787 | | |
| Db | 189675 | GCCTTGACATATATTGGCGGACACGTTTACGG-----TGCGTGGAGCATGACTACC | 189622 | | |
| QY | 788 | TGAATATGTTATGAATAAAACAGCGCGCTTTTCAGATTTAAAGTTGAGACTCATGGAAG | 847 | | |
| Db | 189621 | TGCGGATGTTGATGATAGACCGCGCGCTTGTTCGGCTGACCGTGGTATCATGGAAG | 189562 | | |
| QY | 848 | CGCTGTCTCCTCTCTACA---CCAGCGCAATTCGTTGGTCTCTTCATATAATCTTCGG | 904 | | |
| Db | 189561 | CGTTACGGGAGGCGCCAGATGAGACCGGATCAACATTTGGTGGCGCTTAGCAACCTGCTTG | 189502 | | |
| QY | 905 | GTATTTATTTATCAGATTAGAGATGATTCTGTAATTTGAAGATTTTCCAAATGTCACGG | 964 | | |
| Db | 189501 | GCCTGTCTACCAAGTGGCGGAGCACTATTGAACTTGACCTGACCGACTCGCGATGTCGAAA | 189442 | | |
| QY | 965 | AAAAAGCCTTTGCTGAGGACATTCACAGAGGGGAAGTTATCTTTTCCCATCGTCCAGGCC | 1024 | | |
| Db | 189441 | ACAAGGCCTTTGACAGCATATCACCGAGGGCAATTTCTCTACCCCATATTCACCGGCC | 189382 | | |
| QY | 1025 | TTAACTTCAATAAAACGAAGGTCAAACTGAGCAGACCAATGAATTTCTAAGAAATTTCC | 1084 | | |
| Db | 189381 | TGCAGTACGCCCGTGTCCA---TGACCCAGCGGGCTACGACTTCTCTGTCAGTGTGCTAC | 189325 | | |
| QY | 1085 | TGTTGAGGACAGTGAATAGATATAAACTAAGCTGATTCAAATCTGGAATTCGACA | 1144 | | |
| Db | 189324 | GGCAGAGGACGACAGATATCACCAAGCGCGCTGGTGGGATATCTTGGCGAGCTCA | 189265 | | |
| QY | 1145 | CCAATTCATTTGGCTACACCAAAATTTTATTAATCAATTAAGTGAATATGATAAAA | 1201 | | |
| Db | 189264 | GTGGTTGCTCGGTACACGAGCAGCGGTATCATCGAACTAGCAACATTTGATAAAGA | 189208 | | |
| RESULT 13 | CNS06EHJ | 952 bp | DNA | linear | STS 10-JAN-2001 |
| LOCUS | T3 end of clone AROAA021F04 of library AROAA from strain CBS 732 of | | | | |
| DEFINITION | Zygosaccharomyces rouxii, sequence tagged site. | | | | |
| ACCESSION | AL395117 | | | | |
| VERSION | AL395117.1 | | | | |
| KEYWORDS | STS. | | | | |
| SOURCE | Zygosaccharomyces rouxii | | | | |
| ORGANISM | Zygosaccharomyces rouxii | | | | |
| REFERENCE | Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| AUTHORS | Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces. | | | | |
| | 1 (bases 1 to 952) | | | | |
| | Souciot, J.L., Aigle, M., Artiguenave, F., Blandin, G., | | | | |
| | Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., | | | | |
| | de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorente, B., | | | | |
| | Malpertuy, A., Neuvelise, C., Ozier-Kalogeropoulos, O., Potier, S., | | | | |
| | Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., | | | | |
| | Wincker, P. and Weissenbach, J. | | | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts: 1. A set of | | | | |
| | yeast species for molecular evolution studies | | | | |
| JOURNAL | FEMS Lett. 487 (1), 3-12 (2000) | | | | |
| MEDLINE | 20584711 | | | | |
| PUBMED | 11152876 | | | | |
| REFERENCE | 2 (bases 1 to 952) | | | | |
| AUTHORS | de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B., | | | | |
| | Wincker, P., Artiguenave, F. and Souciot, J. | | | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts: 8. | | | | |
| JOURNAL | Zygosaccharomyces rouxii | | | | |
| MEDLINE | FEMS Lett. 487 (1), 52-55 (2000) | | | | |
| PUBMED | 20584718 | | | | |
| REFERENCE | 3 (bases 1 to 952) | | | | |
| AUTHORS | Genoscope. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |
| | 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : | | | | |
| | seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr) | | | | |
| COMMENT | This STS is part of a random genomic sequencing program of thirteen | | | | |
| | yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces | | | | |
| | exiguus, Saccharomyces servazzii, Kluyveromyces thermotolerans, Kluyveromyces | | | | |
| | Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia | | | | |
| | lactis var. lactis, Kluyveromyces Hansenii var. Hansenii, Pichia sorbitophila, | | | | |
| | angusta, Debaryomyces Hansenii var. Hansenii, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to | | | | |
| | 5 kb were prepared and both extremities were sequenced. See | | | | |
| | keywords for description of this sequence and for the sequence of | | | | |
| | the other extremity of this insert. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..952 | | | | |
| | /organism="Zygosaccharomyces rouxii" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /strain="CBS 732" | | | | |
| | /db_xref="taxon:4956" | | | | |
| | /clone="AROAA021F04" | | | | |
| | /clone_lib="AROAA" | | | | |
| | /note="end : T3" | | | | |
| misc_feature | <154..>270 | | | | |
| | /notes="similar to Saccharomyces cerevisiae ORF YPL070w [| | | | |
| | weak similarity to Vps9p]" | | | | |
| | /evidence="not experimental" | | | | |
| misc_feature | complement (<435..>908) | | | | |
| | /notes="similar to Saccharomyces cerevisiae ORF YPL069c [| | | | |
| | BYS1 ; geranylgeranyl diphosphate synthase]" | | | | |
| | /evidence="not experimental" | | | | |
| ORIGIN | | | | | |
| Query Match | 8.2%; Score 128.6; DB 11; Length 952; | | | | |
| Best Local Similarity | 56.6%; Pred. No. 5.6e-16; | | | | |
| Matches | 303; Conservative 1; Mismatches 215; Indels 16; Gaps 3; | | | | |
| QY | 679 | TTGATTACGATTTTCAACGAGAAATGATCAATCTACATGAGGACAAAGCTTGGATATA | 738 | | |
| Db | 949 | TTTAAATGTTTCTCAATCAGGAAATGATTAAATTTGCATA-GAGCCAAAGGTTAGATATT | 891 | | |
| QY | 739 | TACTGGAGAGACTTCTGCGCTGAAATCATACCTACTCAGGAGATGTTTGAATATGTT | 798 | | |
| Db | 890 | TACTGGCGGATAATTTATTGCGATTTATACCGAGTGAACGCGACGTACTATAATATGTTG | 831 | | |
| QY | 799 | ATGATTAACACAGCGCGCTTTTCAGATTTAAGTTGAGACTCATGGAAGCGCTGTCTCT | 858 | | |
| Db | 830 | ATGAACAAGACTTGGTGGCTTATTAGATTTAAGTTAAGTTAAGTTAAGTTAAGTTAAG | 771 | | |
| QY | 859 | TCCTCACACACCGCCCATTCGTTGGTTCTTTTCAATAAATCTCTGGGTATTTATATCAG | 918 | | |
| Db | 770 | ---GTAGAAATTCACCGCATTCATTAGTACCGCTGAGCAACCTTTTGGGAATAATTTATCAA | 714 | | |
| QY | 919 | ATTAGATGATTTACTTGAATTTTCAAGATTTTCAATGTCACGCAAAAGGCTTGTCT | 978 | | |
| Db | 713 | ATTCGCGATGATTTATCAAAATCTTTTCAATGACAAATGATGATGATGATGATGATGAT | 654 | | |
| QY | 979 | GAGGACATTAACAGAGGGGAAGTATCTTTTCCCATCGTCCACGCCCTTAACTTCACTAAA | 1038 | | |
| Db | 653 | GAAGATATTTCTGAGGTAACTATCTTCCCGATTTATCCAGCGCTTGAAT----- | 602 | | |

QY 1039 ACGAAGGTCAAACTGAGCAACACAAATGAAATTTCTAAGAAATTCCTGTTGAGGCAACT 1098
Db 601 ----ATGTCATCTCAAAACGAACTATTTGCAAGAGCTTACGGTTACGAACAGAT 546
QY 1099 GATAAGATATAAACTAAAGTGAATTCAAATACCTGGAATTCGACACCAATTCATTTGGCC 1158
Db 545 GATTTGAACTGAAAAAAGGCAATTAGACTACCTTAATAACGAAAGTCACCTCCCTAGAG 486
QY 1159 TACACCAAAATTTTATTAATCAATAGTAGATATAGATATAGATATAAAAAATGATATGAAA 1213
Db 485 TACACTTGGGAAAAGCTCGAGGAACCTCTCAAAATTTGGCTAAATTCAGAGAAATATA 431

RESULT 14
AB118242 1130 bp mRNA linear ROD 21-JUL-2004
LOCUS Rattus norvegicus rGGPS1b2 pseudogene mRNA.
DEFINITION AB118242
ACCESSION AB118242
VERSION AB118242.1 GI:50428391
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1
Matsumura, Y., Kuzuguchi, T. and Sagami, H.
Relationship between Intron 4b Splicing of the Rat Geranylgeranyl
Diphosphate Synthase Gene and the Active Enzyme Expression Level
J. Biochem (2004) In press
2 (bases 1 to 1130)
Matsumura, Y., Kuzuguchi, T. and Sagami, H.
Direct Submission
Submitted (26-AUG-2003) Hiroshi Sagami, IMRAM, Tohoku University;
2-1-1 Katsuhira, Aoba, Sendai, Miyagi 980-8577, Japan
(E-mail: yasagami@tagen.tohoku.ac.jp, Tel: 81-22-217-5622,
Fax: 81-22-217-5620)

FEATURES
Location/Qualifiers
1..1130
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/sex="male"
/tissue type="testis"
1..1130
/gene="rGGPS1b2"
/pseudo
1..234
/gene="rGGPS1b2"
235..1062
/gene="rGGPS1b2"
/pseudo
/codon_start=1
1062..1067
/gene="rGGPS1b2"
1063..1089
/gene="rGGPS1b2"

ORIGIN
Query Match 7.8%; Score 121.6; DB 10; Length 1130;
Best Local Similarity 52.7%; Pred. No. 1.5e-14;
Matches 355; Conservative 0; Mismatches 289; Indels 30; Gaps 3;
QY 353 AAAATGAAGCTTGATTCAAACCTTATATCATCATCTTTTGAACCTGGCAAGAACT 412
Db 254 AAGCTGAAGAATTTCTCTAGAGCTTATAAGTACTTCAATTAACAGGTAAACAGG 313
QY 413 TTAGACTAAATTAATAGTTCAAAATTAACAGAGTATGAAATTTGCCAAAGACCGCTGG 472
Db 314 TGAGAACCAACTTTCAGAGCAATTTATCACTGGCTGAAAGTTCCAGAGACCAAGTAC 373
QY 473 CCATAGTTTCGCAATTTGAGCTCTTGCAATAATTCAGCCTTTTAAATCGACGATATAG 532

Db 374 AGATTATCATTTGAGTGAATGTTGCAATAATGCCAGTTTACTCATTTGATGATATTG 433
QY 533 AAGATAATGCTCCCTTGAGAAGGGGACAGACACTTCTCACTTAAATCTTCGGTGTACCCCT 592
Db 434 AAGACAGTTCAAAGCTCCGACGTGTTTCCAGTTGCTCACAGCATCTATGTGTGCCAT 493
QY 593 CCACCTATAAACACCGCAATTTATATGTTTTCAGAGGCCATGCAACTTTGATCGCAGCTAA 652
Db 494 CTGTCAATTAACCTCTGCCAATTTATGTCTTCTTTGGCCTGGAAAAAGTCT----- 544
QY 653 CCACAAAAAGAGCTTTGTATCATATAATTTGATTAACGATTTTCAACGAGAATTGATCAATC 712
Db 545 -----TACACTTGTATCACCAGATGCAGTGAAGCTTTTTCACAGCAGCTTTTGGAAAC 598
QY 713 TACATAGGGGACAAAGCTTTGGATATATATCTGAGAGACTTTCTGCTGAAATCATACCTTA 772
Db 599 TTCAATCAGGACAGAGCCCTAGATATTTACTGAGAGA-----TACCTACACTTGTCTTA 652
QY 773 CTCAGGAGATGTTTGAATATGTTTGAATATAAAGCAGGCGGCGCTTTTCAGATTAAAGT 832
Db 653 CTGAAGAAAGATATAAAGCCATGTTGTTGTCAGAGACAGAGGTGGCTTTGTTGATTAGCGG 712
QY 833 TGAGACTCATGGAAGCGCTGCTCTCTCTCTCACACCCAGGCCATTTCTGTTGGTTCCTTTCA 892
Db 713 TAGGCTCATGCGATTGTTCTCTGACTACAAAGAGATCTAAAGCCACTGCTGG----- 766
QY 893 TAAATCTCTGGGTATTAATTTATCAGATTAGAGATGATTAATCTTGAATTTGAAAGATTTC 952
Db 767 ---ACACACTTTGGGCTCTTCTCCAGATTAGAGATGATTACGCCAATCTACACTCCAAAG 823
QY 953 AAATGTCAGCGAAAGAGCTTTGCTGAGGACATTACAGAGGGAAGTTATCTTTTCCCA 1012
Db 824 AATACAGTGAAGAAACAAAAGTTTCTGTGAAGACCTGACAGAGGAAGTTCTCATTTCTTA 883
QY 1013 TCGTCCACGCCCTT 1026
Db 884 CTATTCATGCCATT 897

RESULT 15
AB118243 1487 bp mRNA linear ROD 21-JUL-2004
LOCUS Rattus norvegicus rGGPS1bdelta pseudogene mRNA.
DEFINITION AB118243
ACCESSION AB118243
VERSION AB118243.1 GI:50428392
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1
Matsumura, Y., Kuzuguchi, T. and Sagami, H.
Relationship between Intron 4b Splicing of the Rat Geranylgeranyl
Diphosphate Synthase Gene and the Active Enzyme Expression Level
J. Biochem (2004) In press
2 (bases 1 to 1487)
Matsumura, Y., Kuzuguchi, T. and Sagami, H.
Direct Submission
Submitted (26-AUG-2003) Hiroshi Sagami, IMRAM, Tohoku University;
2-1-1 Katsuhira, Aoba, Sendai, Miyagi 980-8577, Japan
(E-mail: yasagami@tagen.tohoku.ac.jp, Tel: 81-22-217-5622,
Fax: 81-22-217-5620)

FEATURES
Location/Qualifiers
1..1487
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/sex="male"
/tissue type="testis"
1..1487
/gene="rGGPS1bdelta"

gene

5'UTR /pseudo
1..234 /gene="rGSPS1bdelta"
CDS 235..1038 /gene="rGSPS1bdelta"
/pseudo
3'UTR /codon_start=1
1039..1468 /gene="rGSPS1bdelta"
polyA_signal 1447..1452 /gene="rGSPS1bdelta"
ORIGIN

| | | |
|-----------------------|------|--|
| Query Match | | 7.8%; Score 121.6; DB 10; Length 1487; |
| Best Local Similarity | | 52.7%; Pred. NO. 1.4e-14; |
| Matches | | 355; Conservative 0; Mismatches 289; Indels 30; Gaps 3; |
| QY | 353 | AAAATGAAGAGCTTGATTTCAAAACCTTATAATCACATCCCTTTTGAAACCTGGCAAGAACT 412 |
| Db | 254 | AGCTGAAAGAAATCTCTTAGAGCCTTATAAGTACTTACTTCAATTACCAGGTAAACAGG 313 |
| QY | 413 | TTAGACTAAATTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCACTGG 472 |
| Db | 314 | TGAGAACCAAACTTTCACAGGCATTTAATCACTGGCTGAAAGTTCAGAGAACAAGCTAC 373 |
| QY | 473 | CCATAGTTTCGCAAAATGTTGAGCTCTTGCATRAATCCAGCCTTTTAAATCGACGATATAG 532 |
| Db | 374 | AGATTATCATTTGAAGTGAAGTGTGCTTAATGCCAGTTTACTCATTTGATGATTTG 433 |
| QY | 533 | AGATAATGTCCTCTTGAGAGGGGAGACAGACCACTTCTCACTTAAATCTTCGGTGTACCC 592 |
| Db | 434 | AGACAGTTTCAAGCTCCGAGCTGGTTTCCAGTTGCTCACAGCACTATGGTGTCCCAT 493 |
| QY | 593 | CCACTATAACACCGCAAAATTAATGATTTTCAAGCCATGCAACTTGTATCCAGCTAA 652 |
| Db | 494 | CTGTCAATTAACCTGCGCAATTAATGCTACTTCTTGGCCTGGAAAAAGTCT----- 544 |
| QY | 653 | CCACAAAGAGCCTTTGTTATCATAAATTTGATTACGATTTTCAACGAAGATTTGATCAATC 712 |
| Db | 545 | -----TAACACTTGATCACCCAGATGCGATGAAGCTTTTCAACAGCAGCTTTTGGAAC 598 |
| QY | 713 | TACATAGGGGACAAGGCTTGGATATATACCTGGAGAGACTTTCTGCCCTGAAATCATACCTA 772 |
| Db | 599 | TTCAATCAGGACAAGGCTAGATATTTACTGGAGAGA-----TACCTACACTTGTCTTA 652 |
| QY | 773 | CTCAGGAGATGATTTGAATATGTTATGAATTAATAACAGCGCGGCTTTTCAGATTAACT 832 |
| Db | 653 | CTGAAGAAGAAATATAAAGCCATGGTGTTCAGAGAAGCAGGTGGCTTTTGGATTAGCG 712 |
| QY | 833 | TGAGACTCATGGAAGCGCTCTCTCTCTTCCCTCACACACGCGCATTCGTTGGTTCCCTTCA 892 |
| Db | 713 | TAGGCTCATGCGATGTTGTTCTCTGACTACAAAGAGATCTAAAGCCACTGCTGG----- 766 |
| QY | 893 | TAAATCTTCTGGGTATTATTATATAGATTAGAGATGATTACTTGAATTTGAAAGATTTC 952 |
| Db | 767 | ---ACACACTTGGGCTCTCTTCCAGATTAGAGATGATTACGCCAATCTACACTCCCAAAG 823 |
| QY | 953 | AAATGTCACGCAAAAGGCTTTGCTCAGACATTAACAGGGGAGCTTATCTTTCCCA 1012 |
| Db | 824 | AATACAGTGAACAAAGTTTCTGTGAGAGCCTGACAGAAGGAAGTTCTCATTTCCCTA 883 |
| QY | 1013 | TCGTCCACGCCCTT 1026 |
| Db | 884 | CTATTATGCAATT 897 |

Search completed: October 16, 2004, 23:04:09
Job time : 6610.4 secs